The Importance of Candidate Sex and Partisan Preference Over Time: A Multiday Study of Voter Decision-Making

David J. Andersen
Assistant Professor of Political Science
Iowa State University
dander@iastate.edu

Tessa Ditonto
Assistant Professor of Political Science
Iowa State University
tditonto@iastate.edu

Abstract

Women often face challenges when running for political office, but precisely when and how candidate sex affects voter decision-making is unclear. Using a unique multi-day, high-information experiment, we examine how the presence of women candidates in an election influences subjects' information search, candidate evaluations, and vote decisions. We focus on how the partisan alignment of women candidates (whether they run in the subject's preferred "in-party" vs. their "out-party") matters, and at which point in the campaign gender is most influential. We find that subjects who see in-party women candidates are more open to considering the out-party candidate, seeking out more information about the candidates in the race. Out-party women candidates strengthen subjects' initial partisan preferences, however, leading to less search and higher in-party voting rates. We also find that candidate gender is most influential early in the campaign, and its effects diminish as the campaign progresses.

Keywords: Voter decision-making, Women candidates, Experimental Methods, Candidate evaluation, Dynamic process-tracing environment

- Supplementary material for this article is available in the appendix in the online edition.
- The studies reported within this article were conducted in compliance with relevant laws and were deemed exempt by the Iowa State University Institutional Review Board (IRB# 15-614)
- Data and replication files for this research can be found at: Andersen, David; Ditonto, Tessa, 2019, "Replication Data for: The importance of candidate sex and partisan preference over time: A multiday study of voter decision-making", https://doi.org/10.7910/DVN/SXLXNV, Harvard Dataverse

The Importance of Candidate Sex and Partisan Preference Over Time:

A Multiday Study of Voter Decision-Making

The 2018 midterm elections were momentous for women's political representation. Touted as the new "year of the woman," 2018 saw record numbers of women run for and win office. Despite this progress, women are still under-represented in American politics, comprising less than 25% of the 116th Congress.[1] Prior research has found several facets of the American political system that contribute to women's under-representation, such as the candidate emergence process (Lawless and Fox 2005; 2010) and institutional factors (e.g. Sanbonmatsu 2006; 2010). One possible barrier to women's representation that requires further consideration is the role of candidate sex in the process of voter decision-making (e.g. Ditonto, Hamilton and Redlawsk 2014; Ditonto 2018). Clearly, female candidates are not universally constrained by their sex, as many voters are willing to cast their votes for women on Election Day. On the other hand, gender-based stereotypes, bias, and prejudice continue to shape the experiences of female candidates (e.g. Huddy and Terkildsen 1993; Paul and Smith 2008; Schneider and Bos 2014; Schaffner, MacWilliams and Nteta 2018). It is important, therefore, to understand the contexts in which gender may serve as a barrier to women candidates and when other variables become more influential.

Information Processing Theory (Lau and Redlawsk 2006) suggests that the influence that any piece of information has upon a decision is not necessarily constant, but can change based upon factors such as the point at which it is received, the other information that is available, and how relevant it is to the decision at hand. Thus, candidate sex may matter differently to voters depending on various elements of the larger information environment, and we anticipate that candidate sex may lead voters to seek out information about the two candidates differently, evaluate them differently, and ultimately may also influence their vote choice. For example, candidate sex may matter more in the early period of a campaign, when it does not have to compete with much other substantive information. Sex may then be superseded by substantive information later on, or it may actually guide the ways in which voters learn

[1] Center for American Women and Politics, cawp.rutgers.edu

about and process that later information. Indeed, several previous studies have found that experimental participants conduct different types of information searches for female vs. male candidates (Ditonto, Hamilton and Redlawsk 2014; Ditonto 2017; Ditonto 2018). Also, other experimental research examining campaign information effects has demonstrated that some information, such as campaign ads or negativity, produce an immediate effect but also sees those effects diminish over time (Mitchell 2012, 2013, 2014; Gerber, et al. 2011).

Another important aspect of the information environment during a campaign is partisanship—both of the voter and of the candidates in the race. Most experimental work on candidate gender has measured evaluations of individual candidates in isolation, without considering how the effects of gender may differ based on the sex of the candidates running *both* in a voter's preferred party (their "in-party") and their non-preferred party (their "out-party"). However, all (or at least most) candidates must run against someone else, and voters are likely making/updating their judgments about a particular candidate in relation to the other candidates in the race. If this is true, then a systematic examination of candidate evaluations when the sex of both (in-party and out-party) candidates is considered may also help us to better understand how gender affects not just individual women candidates, but all races in which a woman runs.

It is these variables—time and partisan context—that we seek to incorporate into an analysis of candidate sex and voter decision-making, and we do so by utilizing a novel experimental design. Much of our knowledge about how voter attitudes and behavior affect (and are affected by) candidate sex comes from experimental work, and short, "vignette"-style or survey experiments, in particular (e.g. Huddy and Terkildsen 1993; Holman, Merolla and Zechmeister 2011; Bauer 2015a, b). We have learned much from these studies, as experiments allow researchers to control for the unavoidable complexity in actual campaigns, allowing us to determine which attitudes and behaviors are a result of a candidate's sex, per se, and which can be attributed to other factors. This control comes with diminished external validity, however, and understanding how those effects translate to the real world is tricky.

In an attempt to more closely simulate the "real world" and get a more nuanced look at the process of voter decision-making, we used the Dynamic Process-Tracing Environment (DPTE) to simulate a congressional campaign between two candidates and invited subjects to return each day over a 10-day period to learn new information about the candidates. We randomly manipulated the candidates to appear as either two men or one man and one woman (who could appear in either the subject's in- or out-party), while holding everything else constant. By providing a rich information environment and slowly informing our subjects about the candidates in separate sessions over a relatively long period of time, we were able to provide a more realistic campaign experience than most experimental studies, and were also able to track what subjects learned during the "campaign," when they learned it, and how they reacted to it. This allows us an unprecedented look at how the sex of a candidate mattered to our subjects over the course of a campaign, including initial impressions, information search, the development of candidate evaluations and the final vote choice.

Overall, our results point to an important but previously unobserved connection between candidate sex and over-time voter behavior. We find that candidate sex affects the amount of information that subjects access, how positively they feel toward the candidates they see, and whom they ultimately cast their vote for. Our results suggest that candidate sex seems to matter most at the beginning of a campaign, with the effects of sex waning as our campaign progresses. We also find that candidate sex matters to subjects whether they see a woman candidate in their preferred party or their non-preferred party, and our strongest effects are for subjects who see a woman running in the *other* party. Specifically, subjects who see a woman run in their in-party (and a man in their out-party) look for more information during the campaign than subjects in other conditions, but evaluations of in-party candidates and ultimate vote choice are largely unaffected by in-party candidate sex. On the other hand, subjects who see a woman in their out-party (and a man in their in-party) look for less information during the campaign, evaluate their in-party (male) candidate more favorably than subjects in other conditions, and are less likely to vote for their in-party (female) candidate than are subjects in other conditions. These effects are particularly strong for subjects whose partisan attachments are weak, suggesting that independents and

"leaners" may be most affected by candidate sex in their political attitudes and behaviors. Overall, our results point to a nuanced but important role for candidate sex in voter decision-making.

Gender, Sex and Candidate Evaluation

While many survey-based studies find that bias toward women candidates is not a major obstacle to their election to office (Burrell 1994, Seltzer, Newman and Leighton 1997, Darcy et al 1997, Woods 2000, Dolan 2014), experimental evidence indicates that women candidates may be subject to gender-based stereotypes (e.g. Huddy and Terkildsen 1993, Kahn 1996, Cook 1994; Rosenwasser and Seale 1988). Specifically, voters assign communal personality traits to women; they are viewed as being compassionate, gentle, warm, cautious, trustworthy and emotional (Huddy and Terkildsen 1993; Kahn 1996; Leeper 1991). However, another group of recent studies finds little evidence that feminine stereotypes affect election outcomes (Huddy and Capelos 2002; Dolan 2014; Brooks 2013; Hayes 2011).

Still others suggest that female candidates may not necessarily be stereotyped in traditionally feminine ways (Dolan 2014; Brooks 2013; Hayes 2011; Bauer 2015), but that they may still be evaluated as lacking in the traits necessary for effective leadership (Eagly and Karau 2002; Schneider and Bos 2014; Ditonto 2018) and that this can have negative consequences (Holman, Merolla and Zechmeister 2011). Eagly and Karau (2002), for example, posit that women in leadership positions face prejudice because they are assumed not to have traits that are congruent with leadership roles—things like ambition, competitiveness, and objectivity. Further, Schneider and Bos (2014) find that women who run for office are not stereotyped as women generally, but as women candidates, per se, and are not ascribed positive leadership traits such as competence and expertise.

There is also disagreement on the relationship between partisanship and gender stereotypes. Some scholars have found that partisan considerations are so important that they simply render any possible effects of candidate sex inconsequential (Hayes 2011; Dolan 2014). However, others have found more of an interaction between partisanship and sex. Sanbonmatsu and Dolan (2009), for example, find that voters assign gender stereotypes within both parties, but that this poses a bigger problem for Republicans, and King and Matland (2003) similarly find that Republican women are disadvantaged because they are seen

as more liberal. Finally, candidates who seek to counter gender-based stereotypes may be successful with co-partisans but face backlash from out-party subjects (Krupnikov and Bauer 2014; Bauer 2017).

Impression Formation and Information Processing in Campaigns

It is clear that candidate sex provides voters with a certain amount of information that may be relevant to their decision-making. What remains unclear is the extent to which voters consider gender-based information in their vote calculus given the other politically relevant information that is often available to them during a campaign. In an attempt to shed light on this question, we employ Information Processing Theory (Lau and Redlawsk 2006). We treat candidate evaluation and voting as information-processing tasks and the act of casting a vote as only the last step in an often long, complex process.

According to this paradigm, voters do not come into political campaigns as "blank slates," but instead bring existing partisan attitudes, stereotypes and beliefs with them (Taber and Lodge 2013). Much of the early information that voters then acquire in a campaign likely comes from rapid and heuristic-based "System 1" processing, which includes things gender, race, and other attributes that can be discerned via someone's physical appearance (Fiske 1998; Lau and Redlawsk 2006). Over the course of a campaign, voters then search for and encounter substantive information about candidates, which updates their earlier inferences (Fiske and Neuberg 1990; Fiske and Neuberg 1999; Taber and Lodge 2013).

This model suggests that candidate characteristics like sex can, and likely do, affect vote choice in two ways. First, they may alter initial impressions. Voters hold preconceptions about the parties and their candidates, but when voters first meet a particular candidate, they can update these predispositions, using visual cues to help them make assumptions about what that particular candidate is like. Second, cues can then also influence information search and processing. For example, voters who have strong party preferences, but meet a candidate who does not match their expectations (i.e. a woman instead of a man) may alter their preferences for that candidate *and then* seek out more information about that candidate, to make sure that they are satisfactory.

There is evidence that candidate sex leads experimental subjects to seek out both more information and different kinds of information about female candidates, when compared to male candidates, and that they react differently to the information they encounter when it is about a woman vs. a man. Ditonto, Hamilton and Redlawsk (2014) find that subjects access more information overall, and more information about a candidate's stance on "compassion issues," as well as competence and qualifications, when that candidate is a woman compared to a man. Ditonto (2017; 2018) further finds that subjects care more about competence-related information when it is about a female candidate. They both evaluate a female candidate less favorably and are less likely to vote for her when she is portrayed as less competent in the information available about her, whereas men associated with similarly unflattering information are not affected.

Further, we expect that candidate gender matters both for candidates in a voter's in-party and in their out-party. The process of coming to a vote decision involves encountering and weighing information about one candidate in relation to what one knows about the others in the race. Asking subjects to evaluate candidates in a simulated campaign is inherently asking them to complete a relational task—judgments about one candidate likely affect evaluations of the other, whether they realize it or not. In a general election context, those relational judgments are likely to be made through the lens of partisanship, which is almost universally acknowledged as the most influential cue in American politics (e.g. Campbell et al 1960; Zaller 1992; Bartels 2000; Goren, Federico and Kittilson 2009).

Uncertainty about the candidate in one's own party can lead a voter to rely less heavily on partisanship to make their decision in a political contest (Basinger and Lavine 2005; Lavine, Johnston and Steenbergen 2012). This may lead them to consider factors other than partisanship when forming judgments about candidates and cause them to seek out more information about the candidates in the race (Marcus and Mackuen 1993; 2000; Ditonto et al 2014). It may also ultimately make them more likely to vote for an out-party alternative than they would be in other circumstances. This may be especially true for weak partisans (Basinger and Lavine 2005; Lavine et al 2012).

Because female candidates are often subject to stereotypes that paint them as less suitable for public office, seeing an in-party woman running for office may lead subjects to feel less certain about their in-party candidate, leading voters to more seriously consider an out-party candidate (provided he is a man). At the same time, seeing a woman running in one's out-party may both reinforce positive evaluations of a (male) in-party candidate and activate negative stereotypes related to women's competence and unsuitability for political office, making an out-party female candidate a particularly unattractive political choice.

Hypotheses and Method

We expect to find differences in how people search for information about the candidates, evaluate the two candidates throughout the campaign, and ultimately choose to vote, dependent upon the overall gender composition of the election scenario they view. Voters who see two men challenging each other should rely primarily on partisanship cues to make their evaluations and vote decisions, and their information search should reflect this, as well. They should prefer their in-party candidate to their out-party candidate to a large degree and be much more likely to vote for their in-party candidate over their out-party candidate. This is typically what voters encounter during an American election, and is the baseline of political behavior.

When voters see a woman candidate run, however, this alters how they react to the candidates, dependent upon whether that woman is running in their preferred party or the opposing party. While more women are running and being elected to office than ever before, women candidates are simply still not the "typical" candidate, and lingering stereotypes about their incompatibility with leadership roles (e.g. Eagly and Karau 2002) persist. This may make voters question whether their typically-dominant partisan predispositions still hold true and may jolt them out of System 1 thinking and trigger a System 2 information search (Chaiken 1980; Petty and Cacciopo 1981; Chaiken and Trope 1999; Bargh 1999).

Whereas subjects who view two male candidates should rely primarily on partisanship cues:

H1: Subjects who view a woman candidate running in their in-party should be less sure about their partisan predispositions, and therefore more likely to consider their out-party candidate as a viable option. They will:

H1a: Search for more information about the candidates overall, and particularly more information about the in-party candidate.

H1b: Have weaker preferences for their in-party candidate over the out-party candidate.

H1c: Be more likely to defect from their in-party and vote for the out-party candidate

H2: Subjects who view a woman candidate in their out-party should be more certain about their partisan predispositions, and therefore be less likely to consider their out-party candidate. They will:

H2a: Search for less information about the candidates overall, and particularly less information about the out-party candidate.

H2b: Have stronger preferences for their in-party candidate over the out-party candidate.

H2c: Be less likely to defect from their in-party and vote for the out-party candidate.

H3: Observed candidate gender effects should be stronger at the beginning of the campaign, and attenuate over time.

H3a: Information search differences should appear in the beginning of the campaign, but as voters accumulate information about the candidates over time, those differences should disappear.

H3b: Gender-based effects on candidate preference should be strongest at the beginning of the campaign, but decrease over time.

Method

In this paper, we use the Dynamic Process Tracing Environment (DPTE),[2] which allows us to simulate a political campaign with greater complexity than traditional survey experiments. We invent

[2]See http://www.processtracing.org for the Dynamic Process Tracing Environment (DPTE) software and user guide. Any researcher may request access to the system for research purposes by clicking on the appropriate link on the website.

candidates and experimentally manipulate their gender, while allowing subjects to choose from a broad range of candidate information in an environment that mimics the dynamic nature of a real-world political campaign (Lau and Redlawsk 2006). This method provides us with a unique opportunity to leverage the control provided by experiments while also more closely simulating the information-rich and constantly changing nature of a real campaign environment, thus improving external validity. Embedding experimental manipulations within a more complex information environment has been shown to reduce treatment effects and produce results more closely approximating real world effects (Andersen and Ditonto, 2018).

Most DPTE studies present subjects with candidate information over a comparatively long timeframe, typically 15-30 minutes, permitting researchers to study the process of how voters learn about candidates and arrive at vote decisions (e.g. Lau and Redlawsk 1997, 2001, 2006; Lau et al 2017). However, even these types of studies make it difficult to see which information ultimately matters and *when* it's making a difference, as well as how evaluations of candidates may evolve over time. This study (conducted in November of 2015) utilizes DPTE to present information over a longer time frame – 10 weekdays.[3] While this is not the first experimental study to monitor subject performance over time (see Mitchell 2012, 2013, 2014; Gerber et al 2011; Druckman, Fein and Leeper, 2012; and Chong and Druckman, 2013), this is the first study to use DPTE to simulate a campaign over multiple days, creating an information environment far more complex and realistic than anything previously simulated.

In real campaigns, candidates slowly release issue positions, focusing on general themes, reacting to current events and building a campaign that shapes their candidacy. Candidate evaluations evolve slowly as voters hear more about the candidates and form impressions of them. In this study, we attempt to mimic that process. Each weekday for two weeks, subjects were able to sign back into the study and learn new information about the candidates, while we recorded what they chose to view, and how their

[3] The complete study can be viewed by going to: https://bit.ly/2KOmBfs (pop-up blockers must be turned off for the study to load, as it will automatically generate a pop-up window that will fill the screen)

evaluations of the candidates changed. By manipulating the sex of the candidates that subjects saw, we are able to monitor when, where and how candidate sex influenced subject reaction to and perception of the candidates over time. Unlike many previous experiments, this presents our treatment not as a single one-shot stimulus, but as a recurring, stable feature of our study. That is, our manipulation presented our candidates as either a man or a woman and monitored how that mattered over time to our subjects. This provides us the ability to see whether any effects of candidate sex persist over time, attenuate, or completely disappear.

Our sex manipulation allowed one of the two major-party candidates in a simulated congressional election to appear as a woman for our subjects, while all other information remained the same. On the first day, subjects were randomly assigned to see either a Democratic woman candidate, a Republican woman, or two men as candidates.[4] For the remainder of the study, all subjects saw identical information about the candidates, and continued to see those candidates as either a man or a woman, as depicted through a picture[5] of the candidate that appeared on each information box, as well as gendered pronouns in the information box text, and gendered names (Brent/Brenda Evans (D) and Nick/Nicole Robinson (R)).

Sample and Procedure

We recruited 400 subjects from Amazon's Mechanical Turk to participate within this study. On the first day, 383 of those subjects successfully completed the approximately 45-minute long session and received a link that allowed them to return to the study for the remaining nine days. They were paid $4 for that first day's participation. On each day that they returned to the study for the next eight days, they were

[4] We did not include a woman-woman condition because such contests are still exceedingly rare. In 2016 only 15 of the 435 congressional campaigns featured two women running against each other. Given the limits on the size of our sample we opted to produce gender dynamics that are more common in American elections.

[5] See appendix Figure A3 for those pictures.

paid an additional $1, until the final day of the campaign, when they were paid an additional $4 for logging in and casting their vote. Days two through nine were significantly shorter, and only about 3-5 minutes to complete. Subjects who participated on each and every day were thus able to earn $16 in total. We suffered a large attrition rate at the beginning of the study, with about one-fifth of subjects failing to return after the first day. Of those who did return however, most returned each and every day thereafter, with nearly two-thirds completing at least 9 of the 10 days. The attrition of subjects was balanced across groups, and does not seem to have affected our results in any way. More information on the attrition rates and composition of our sample can be found in the Appendix.

Table 1. Information available about each candidate

Attributes	Availability									
	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9	Day 10
Abortion Policy	0	4	2	1	1	0	0	0	1	1
Iran Policy	0	0	0	0	0	0	0	4	2	1
Crime Policy	0	0	0	0	0	0	0	0	4	1
Economic Philosophy	0	0	4	2	1	1	0	0	0	1
Editorial About	0	0	0	4	2	1	1	0	0	1
Education	2	1	0	0	0	0	0	0	0	1
Education Policy	0	0	0	0	0	4	2	1	1	1
Energy Policy	0	0	0	0	4	2	1	1	0	1
Family	2	1	0	1	0	0	0	0	0	1
Global Warming Policy	0	0	0	0	0	4	2	1	1	1
Gun Control Policy	0	4	2	1	1	0	0	0	1	1
Healthcare Policy	0	0	0	0	0	0	4	2	1	1
Immigration	0	0	0	0	0	0	4	2	1	1
Jobs Policy	0	0	0	4	2	1	1	0	0	1
Political Experience	3	2	1	0	0	0	0	0	0	1
Religion	3	1	1	0	0	0	0	0	0	1
Tax Policy	0	0	4	2	1	1	0	0	1	1
Terrorism Policy	0	0	0	0	4	0	0	0	0	0
Social Philosophy	3	2	1	0	0	0	0	0	1	1
Defense Policy	0	0	0	0	0	0	0	4	2	1
Items per Candidate	13	15	15	15	16	14	15	15	16	19
Total	26	30	30	30	32	28	30	30	32	38

On each day, subjects were presented with a new set of information about the two candidates. The same types and amounts of information were available about both candidates on each day, and we clustered information together to form thematic days (i.e. on day three candidates introduced their economic ideas by "talking about" their economic philosophies and tax policies). Each day presented two new types of information about the candidates that they "emphasized" (four information boxes appeared containing the information), while they "continued" to discuss the previous day's information to a lesser

degree (one or two boxes). In total, subjects could learn 20 different attributes about the candidates, spread over the 10 days. Table 1, above, shows the specific issues that were available on each day. A more detailed discussion of how the study presented information about the candidates is in the Appendix.

The benefit of a simulation such as this is that we are able to craft exactly the experimental conditions we wish to examine, precisely monitor subject behavior and take numerous measurements of subject opinions throughout the study, creating a robust series of measures with which to test our hypotheses. The negative consequence is that the costs of fielding such a simulation are relatively high – in this case approximately $8,000 to recruit just 400 subjects (relying on ethical pay rates as discussed in Andersen and Lau 2018). Because of the relatively small sample size our ability to draw out statistical significance is limited. The nature of our results however, taking multiple measurements over the course of 10 days, permits us to recognize patterns in the data that suggest continuity in our findings. This consistency, while not statistical significance, should provide similar levels of heightened confidence that our results are in fact valid and not spurious. For that reason, we will reference some findings that lack standard levels of statistical significance ($p < .05$ using a two-tailed test) but are consistent throughout our results.

Results

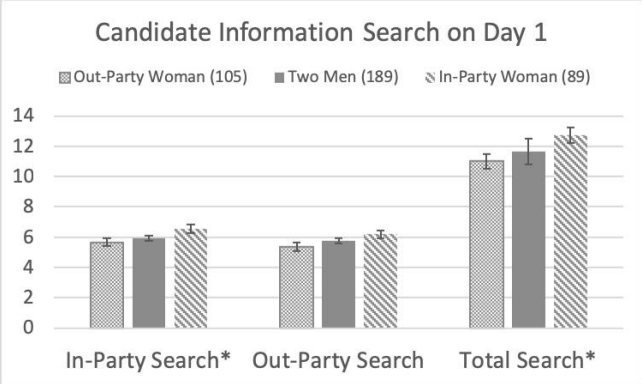
Day 1 Results

The unique design of our campaign simulation allows us to take daily "snapshots" of how voters reacted to the candidates. Since most experimental research on candidate sex essentially takes a single snapshot, we first present the results from our first day, allowing our study to replicate and compare to existing scholarship. On the first day, subjects "met" the candidates in a format quite similar to how subjects learn about candidates in more typical studies, such as vignette-style survey experiments (e.g. Huddy and Terkildsen 1993a, b). That is, subjects saw a picture of each of the candidates and then were able to learn some basic background information about them, including their education, family, political experience, religion and social philosophy. This mimics how most experiments examine candidate gender (gathering evaluations after brief exposure with very limited information). For us, it serves as a baseline

and shows how our subjects initially reacted to seeing men vs. women candidates in both their in- and out-parties. Unlike previous studies, we can then examine how these initial impressions changed over time.

On day 1, we find that our gender manipulations produced significant effects both in how subjects initially sought out information about the candidates they saw and in how they subsequently evaluated those candidates. Figure 1, below, shows the results of oneway ANOVAs on the amount of information subjects opened about the candidates on the first day, along with their plotted standard errors. Subjects consistently sought more information about their in-party[6] candidate over their out-party candidate, but the magnitude of that difference varied based on the gender of the candidates.

Figure 1. Information search for the candidates on day 1



Subjects who saw two male candidates looked at an average of 5.92 pieces of information about their in-party candidate, and 5.76 information boxes about their out-party candidate, for a total of around 11.67 items overall.[7] Since the vast majority of congressional elections involve two male candidates, we can assume that this is how voters might typically learn about candidates during a congressional race.

[6] All subjects were categorized into a binary variable indicating whether they were closer to the Democratic or Republican Party. Pure Independents were sorted based upon their ratings of the two parties on the feeling thermometers in the preliminary questionnaire.

[7] The table showing the information for Figure 1 is available in Table A2 in the Appendix.

When subjects saw an in-party woman, however, they generally looked for significantly more information about both in-party and out-party candidates (6.54 items for the in-party candidate, and 6.19 items for the out-party, 12.73 items overall). [8] This supports our expectations in hypothesis H1a. Because the presence of an in-party female candidate seems to induce subjects to gather more information, overall, we believe it suggests that they are less certain of their choice than those who see two male candidates. We interpret this to mean that a vote for an in-party woman is less of a foregone conclusion for subjects than is a vote for an in-party man. At the same time, an out-party man running against a woman (even in one's own party) is treated as a viable alternative to a greater extent than an out-party man competing against an in-party man.

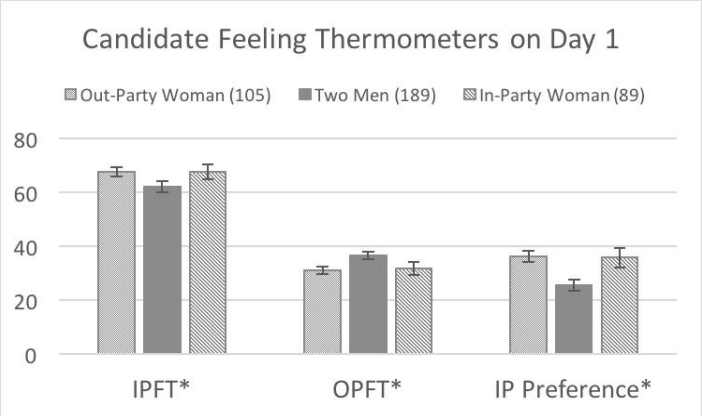
Support for this idea also comes from examining what happened when subjects saw a male in-party candidate running alongside an out-party woman, as subjects in this condition looked for significantly *less* information about both in- and out-party candidates (5.66 vs. 5.35 items, respectively and 11.01 items in total). This supports our expectations in hypothesis 2a, and suggests that these subjects were more comfortable relying upon partisan cues to inform their judgments, and did not feel compelled to seek out as much substantive information about the candidates. Whereas the presence of an in-party female candidate increases information-seeking behavior, then, seeing an out-party woman seemed to reduce the amount of time and effort subjects spent learning about their choices. In sum, candidate gender seems to interact with partisanship such that it affects how comfortable subjects are relying on their initial partisan attitudes and leads them to adjust their information search accordingly.

Candidate gender also significantly affected how subjects felt about the candidates they evaluated on Day 1. When subjects saw two male candidates, on average they rated their in-party candidate at 62.28 points on the 100-point feeling thermometer, while rating the out-party candidate at just 36.66 points (see Figure 2, below and Table A8 in the Appendix). This gave the in-party candidate a 25.62-point edge, on

[8] Significance in the figures is indicated with a single star (*). This signifies that the one-way ANOVA is significant at $p. < .050$ using a two-tailed test. Tables not included in the paper are in the Appendix.

average (In-Party Feeling Thermometer – Out-Party Feeling Thermometer). However, when the out-party candidate was a woman (and the in-party candidate a man), the average in-party evaluation was 67.74, while the out-party score was a 31.27, resulting in a preference gap of 36.48. Since the available candidate information was identical, candidate gender must be driving these results. This significantly larger preference for the in-party man when the out-party candidate is a woman supports our expectation in hypothesis 2b.

Figure 2. Average Feeling Thermometer scores for the candidates on day 1



Interestingly, when the in-party candidate appeared as a woman, subjects also liked her significantly more, and the out-party male significantly less, than when they saw a male candidate in each party (in-party score = 67.67, out-party score = 31.27, and difference score = 35.4). This is contrary to our expectation in hypothesis 1b. Seeing a woman run in either party, then, seemed to lead subjects to like their in-party candidate more and their out-party candidate less than seeing two men run against each other, at least initially.

Because we are interested in the entire process of voter decision-making, one final way to examine how subjects evaluated the candidates on the first day is by controlling for subject partisanship and prior attitudes to isolate the effects of the candidate gender manipulation. Voters bring many prior attitudes and preferences to their evaluations of candidates, especially when it comes to partisanship. Particularly in today's polarized political environment, self-described partisanship only captures some of the preexisting attitudes towards the two parties and their candidates. Many subjects who declare

themselves to have "weak" or no party ties still report substantial preferences for one party over the other and hostility toward one's opposing party is a stronger predictor of political behavior than positive feelings toward one's in-party (Iyengar and Krupenkin 2018). We account for this by including a measure of subjects' attitudes toward typical political candidates from each of the two major parties *before* meeting the actual candidates in the study. In order to calculate this "initial partisan preference score," we collected feeling thermometer scores for "most Democrats" and "most Republicans" then took the difference between a subject's in-party rating and out-party rating. This provides us with a more nuanced measure of existing attitudes than simply relying on declared partisanship.

Table 2, below, shows the results of a Generalized Method of Moments using the Day 1 candidate preference score as the dependent variable, and including the initial partisan preference score discussed above, the traditional strength of party ID measure, and a dummy variable for "Democrat" as predictors. While the results are nearly identical to a standard OLS regression, the GMM is a better fit for dynamic panel data models such as this 10-day study (see Wawro 2002).

Table 2. Generalized Method of Moments for Day 1 In-Party candidate preference scores

	Day 1 (383)
Democrat	14.724*** (3.301)
Strength of PID	0.936 (1.620)
Initial In-Party Preference	0.472*** (0.059)
In-Party Woman	6.807* (3.470)
Out-Party Woman	8.001* (3.088)
Constant	0.391 (3.375)

* - p < .050, ** - p < .010, *** - p < .001

We find that self-described Democrats like their in-party candidate more, and initial party preference also has a positive and significant effect on candidate preference (though strength of party ID does not). In terms of candidate gender, seeing either an in-party or out-party woman candidate

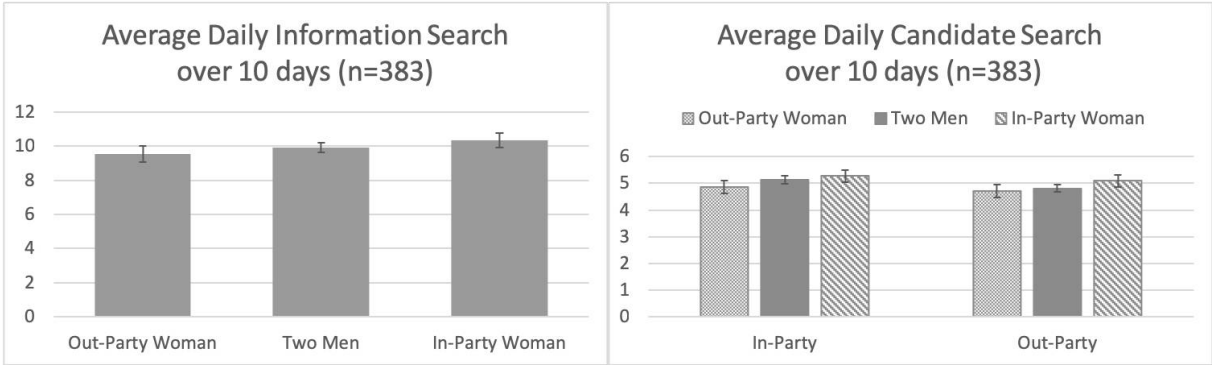
significantly boosted subject's preference for their in-party candidate, even controlling for these priors. This mimics what we found in our simpler analysis above, though it is interesting to note that the magnitude of change is greater for seeing an out-party woman than an in-party woman (8 vs. 6.8 points).

In sum, the Day 1 results provide some clear support for our hypotheses and show how candidate gender affected subject behavior at this early stage of our study. Subjects who saw a woman candidate (in either their in- or out-party) had significantly different information search patterns. They also both liked their in-party candidate more and their out-party candidate less than subjects who saw two male candidates. Clearly, the appearance of a woman candidate in the race mattered, and affected how subjects approached and evaluated the two candidates at the beginning of our "campaign." What other studies have been unable to assess up until now, though, is how these initial evaluations changed over time.

Daily Information Search Over Time

The 10-day duration of this design permits us to see exactly how our subjects chose to learn about the two candidates and to determine whether any differences in search patterns attributable to candidate gender persisted over time. Here, we examine how the average amount of information looked at each day was affected by candidate gender, using a method similar to our Day 1 analysis.[9]

Figure 3. Total Information Search, by Candidate Gender



[9] We also conducted an analysis of search for different types of information. These findings are available in Tables A6, A7a and A7b in the online appendix.

We find no significant differences by candidate gender when we look at total information search over the entire experiment at once, however, we *do* find consistent patterns throughout the campaign. In the first panel of Figure 3[10], we can see that there is an apparent, though not-significant, difference in information search in which subjects examined more information when they saw an in-party female candidate and less information when they saw an out-party female candidate, relative to when they saw two male candidates. This mirrors the significant effects we found on Day 1 and supports our expectation in Hypothesis 3a. If significant differences in information search occur at the beginning of the study, but then fade away to non-existent, or simply not-significant differences later on, the overall result would still demonstrate the initial pattern, but not achieve statistical significance, which is what we observe here. When we break out information search by the candidate being examined (Average Candidate Search, over 10 Days, Figure 3), the pattern of results repeats, with subjects viewing more information about both in-party and out-party candidates when the in-party candidate is a woman, and less information about both when the out-party candidate is a woman. Again, these are not statistically significant findings, but are instructive, given the consistent pattern.

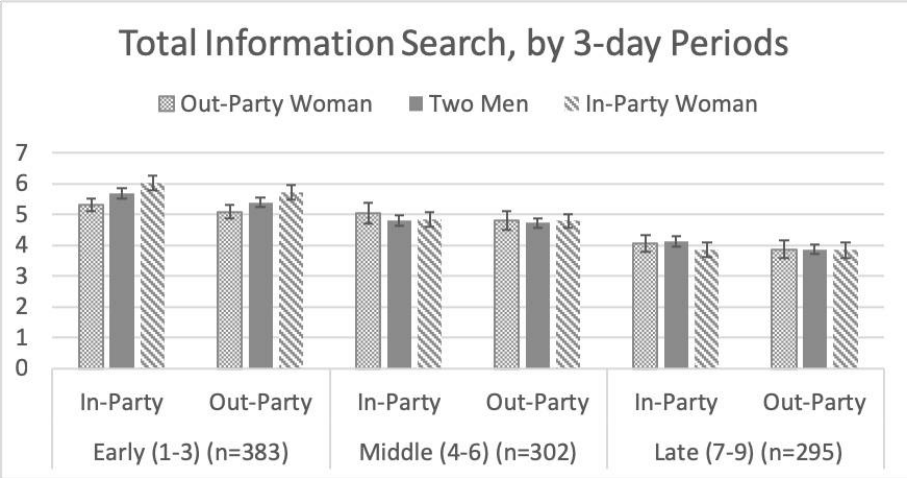
We conduct one final analysis of total information search over the course of the study by dividing the campaign into three "periods" and looking at average information search in the early (days 1,2 and 3), middle (days 4,5 and 6) and late (days 7,8 and 9) periods of the campaign. If the effects of candidate sex diminish over time, we should see the pattern of results of day 1 appear generally in the early part of the campaign, and then disappear subsequently in the latter parts of the campaign. This is what we observe.

Several interesting patterns emerge in Figure 4, below, showing the results of oneway ANOVAs of the average items opened for each candidate and their standard errors, by period. First, it becomes apparent that subjects tended to reduce their information search as the campaigns wore on. In each period, a nearly identical amount of information was available for subjects to view about the candidates and new information appeared each day. However, the information search for both candidates, whether male or

[10] The tables that correspond to Figures 5 and 6 are available in the Online Appendix (Tables A3 and A4)

female, tended to decrease from the Early to the Middle period, and then decreased again from Middle to Late. Subjects clearly attempted to learn a great deal about the candidates early on in the campaign, but grew less active over time. This is likely because their opinions about the candidates became more certain, and the marginal utility of learning new information quickly decreased (we return to this later in our discussion of feeling thermometers).

Figure 4. Periods of Candidate Information Search, by Candidate Gender



Importantly for our gender analysis, we find that the day 1 search pattern (that out-party women decreased overall information search and in-party women increased search) appeared prominently in the early stages of the campaign, but then flattened out in later sections of the campaign. Though these differences in the early stage do not reach conventional standards of statistical significance ($p < .094$ for In-Party; $p < .134$ for Out-Party), they are consistent with the predictions of hypothesis 3a. It appears that the effects of gender on information search are most influential early in the campaign, and disappear over time.[11] Subjects may begin the campaign with skepticism about women candidates because of stereotypes they hold, but over time they may then replace those stereotypes with actual information that dispels any

[11] Figure A2 in the Appendix further reduces this analysis to the daily level, where it is clearly apparent that Days 1 and 3 drive the results in the Early part of the campaign.

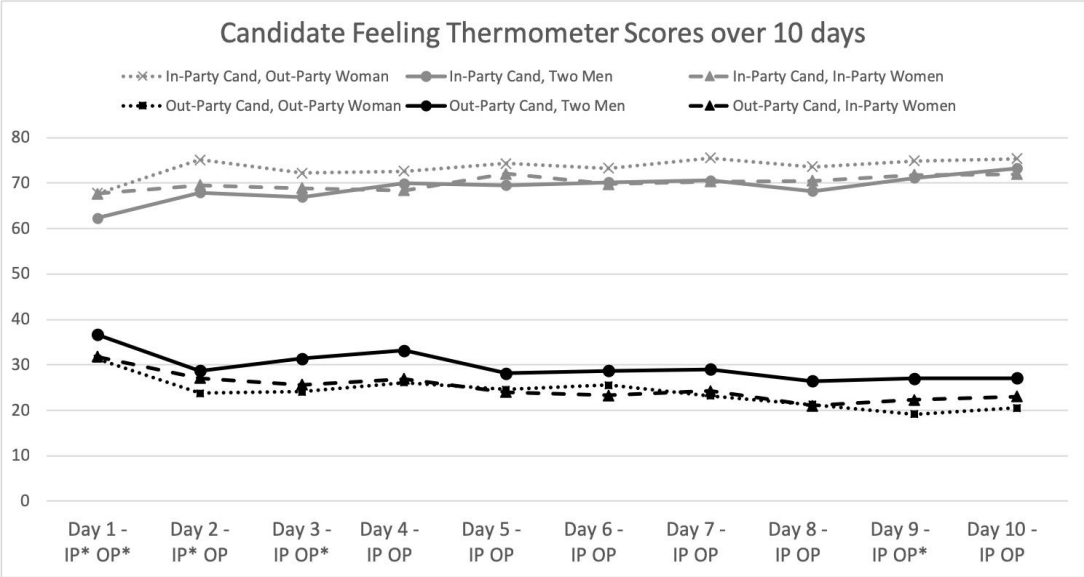
worries that may come with them. If this is the case, we would also expect to see a similar reduction in the influence of gender on candidate feeling thermometer scores.

Daily Candidate Preference Over Time

To track how candidate evaluations changed over time, we first reproduce the one-way ANOVAs that we conducted for candidate feeling thermometers on Day 1, before moving to a more nuanced analysis. Figure 5, below, plots the average feeling thermometer score for subjects' in-party (grey lines, on top) and out-party (black lines, on bottom) candidates, based on the gender composition of the race they saw (significance for each candidate denoted by an "*" on the horizontal axis, where IP is in-party and OP is out-party).

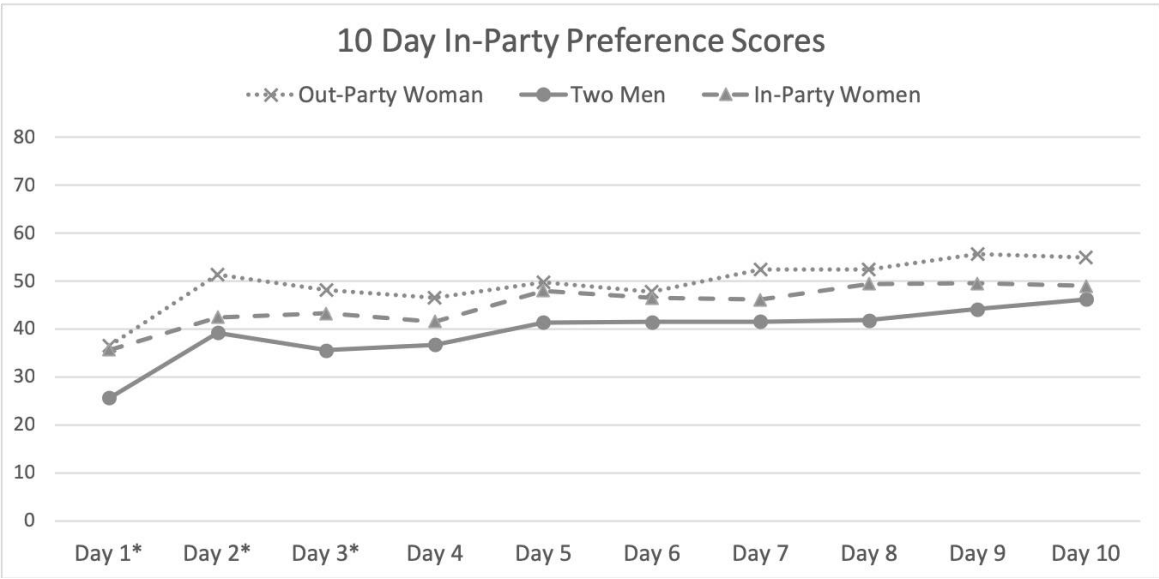
The in-party candidates were always rated higher, and subjects consistently moved towards liking their in-party candidate more, and their out-party candidate less over time. We also find that the initial differences that were observed on Day 1 persisted over time. However, similar to our findings related to information search, candidate gender only had significant effects for a short time. In-party evaluation differences by gender only lasted for the first two days, while out-party differences by gender are significant only on the first, third and ninth day. As with information search, it appears that the influence of candidate gender wanes over time.

Figure 5. Candidate Feeling Thermometers over 10 days, by Candidate Gender



We similarly plot the 10-day candidate preference scores in Figure 6 below (in-party candidate scores– out-party candidate scores). Unsurprisingly, subjects tended to increase their preference for their in-party candidate over time, as they learned more about them and their opponent. On Day 1, we found significant differences in candidate preference based upon the gender of the two candidates. Those differences again persisted throughout the study, but were only significant for the first three days. On each day, subjects reported preferring their in-party candidate most when the out-party candidate was a woman, and least when both candidates were men, but the magnitude of these differences decreases over time.

Figure 6. In-Party Candidate Preference over 10 days, by Candidate Gender



Finally, we generate a generalized method of moments model using the daily preference scores for each day of the campaign. Due to the size of the table required to display these results, we present them in two tables below. Table 3a shows the first week's results, and Table 3b shows the results for week two (below, on page 23). Several patterns emerge looking at these tables. First, the most influential and important predictor of candidate preference throughout is the initial party preference score. While perhaps not surprising, what is surprising is that the magnitude of this variable strengthens over time. Our subjects clearly came into the study with existing preferences, and those preferences were strengthened over time as subjects learned about the candidates in the race.

Table 3a. Generalized method of moments estimation on Week 1 In-Party Candidate Preference Scores

Variable	Day 1	Day 2	Day 3	Day 4	Day 5
Democrat	14.724***	23.498***	10.707**	21.480***	22.073***
	(3.301)	(3.995)	(4.135)	(4.895)	(4.806)
Strength of PID	0.936	0.609	1.200	0.457	-0.151
	(1.620)	(2.142)	(2.104)	(2.426)	(2.548)
Initial In-Party Preference	0.472***	0.588***	0.639***	0.623***	0.617***
	(0.059)	(0.066)	(0.073)	(0.093)	(0.087)
In-Party Woman	6.807*	1.365	4.743	3.040	3.489
	(3.470)	(4.007)	(4.326)	(4.525)	(4.522)
Out-Party Woman	8.001*	7.404*	8.020*	6.636	4.165
	(3.088)	(3.257)	(3.477)	(3.720)	(3.720)
Constant	0.391	4.986	6.099	1.538	7.042
	(3.375)	(4.120)	(4.426)	(4.993)	(5.149)
N	383	272	274	265	243

* - p < .050, ** - p < .010, *** - p < .001

Table 3b. Generalized method of moments estimation on Week 2 In-Party Candidate Preference Scores

Variable	Day 6	Day 7	Day 8	Day 9	Day 10
Democrat	18.913***	10.720**	5.430	16.145***	13.366**
	(5.171)	(4.083)	(4.500)	(4.533)	(4.587)
Strength of PID	0.124	1.758	1.950	-0.494	0.882
	(2.449)	(2.517)	(2.667)	(2.490)	(2.553)
Initial In-Party Preference	0.708***	0.793***	0.826***	0.785***	0.756***
	(0.093)	(0.081)	(0.080)	(0.077)	(0.081)
In-Party Woman	2.927	2.111	2.156	1.665	0.555
	(4.507)	(4.444)	(4.448)	(4.588)	(4.681)
Out-Party Woman	4.032	4.561	2.561	4.861	3.779
	(3.882)	(3.443)	(3.650)	(3.456)	(3.516)
Constant	6.342	7.028	10.335	10.617*	12.095*
	(4.478)	(4.573)	(5.395)	(4.781)	(4.805)
N	249	248	256	256	278

* - p < .050, ** - p < .010, *** - p < .001

Second, the Democrats in our sample clearly preferred the Democratic candidate much more than our Republicans liked the Republican candidate. On almost all days, our Democrats scored more than 10 points higher in candidate preference. This is an unexpected result that is likely tied to the nature of our experimental design, and we would not necessarily expect this to be a generalizable result, as there is no evidence that Democrats in general tend to prefer their party's candidates more than Republicans prefer theirs.[12] It could be because we were better at portraying traditional Democratic candidates than traditional Republican candidates, or that the manner in which subjects learned about the candidates (reading issue positions and background information) was a style preferred by Democrats over Republicans. This is an area for future study, and we will discuss this more in our conclusions.

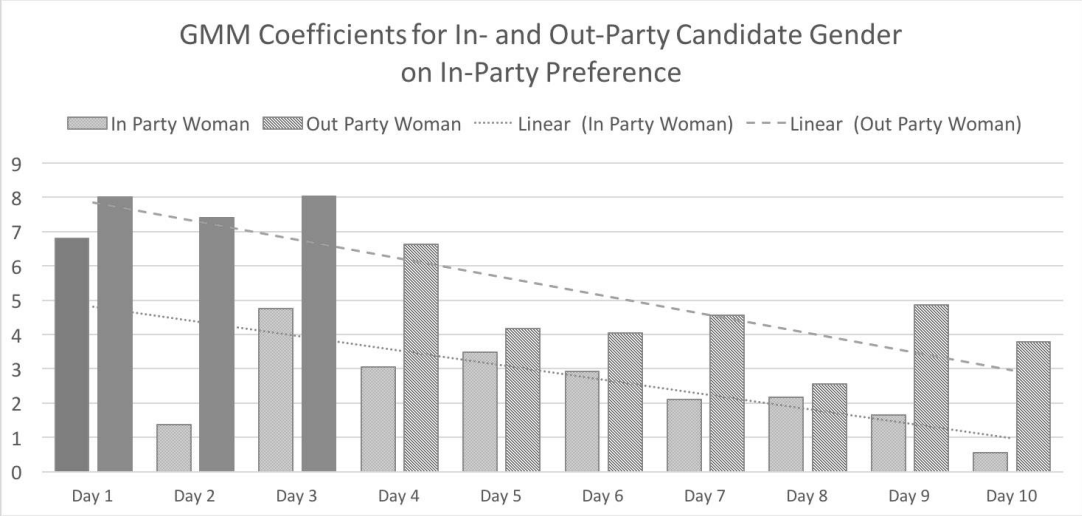
Finally, and most directly related to this paper, we find that the main effects of candidate gender upon candidate preference decline throughout the ten days. On the first day, seeing an in-party woman or an out-party woman significantly strengthened subjects' preference for their own in-party candidate by 6.8-points and 8-points, respectively. But after that, the coefficients for these variables decline fairly steadily, both by magnitude and significance levels, with In-party women dropping below statistical significance on Day 2, and Out-party women dropping below statistical significance on Day 4. Given the relatively small nature of our sample, and the fact that the sample suffered attrition throughout the study (though most of our attrition occurred immediately after Day 1), some of the decline in statistical significance is likely related to a small N. However, the magnitude of the coefficient also steadily and consistently decreases, strengthening our confidence that we are measuring a decrease in the influence of the role of gender throughout the study.

This is perhaps most easily seen graphically. Figure 7, below, charts the coefficients for the in-party and out-party women candidate variables over time. The solid columns represent statistically

[12] Though, at the time this study was fielded, in November 2015, the 2016 presidential primaries were active and the Republican Party was divided between 17 candidates. It is possible that this is what we are observing in this study.

significant findings, while the shaded columns represent non-significant findings. Only four coefficients are significant, one for the presence of an in-party woman candidate on day 1, and three for the presence of an out-party woman candidate on days 1, 2 and 3. A fitted regression line is superimposed over both sets of columns, illustrating the downward trend of both sets of coefficients. Clearly, day-after-day, the measured role of the influence of candidate gender on candidate preference got smaller and smaller.

Figure 7. GMM Coefficients for In- and Out-Party Candidate Gender on In-Party Candidate Preference



There are only two possibilities that we can think of to explain why gender becomes less influential over the course of the campaign – time and information. It is possible that, given the extra time to consider candidate gender, subjects simply overcame any initial skepticism of women candidates. That seems unlikely, as time alone is not likely to aid people in recognizing and correcting an implicit bias. The alternative is that information – in this case in the form of issue positions and background information about the candidates gained through the dynamic information boards – allowed subjects to replace proxy-beliefs that they held about what women candidates would be like with actual information about what these candidates were really like. In learning about the candidates, our subjects overcame existing biases by encountering information (Fiske and Neuberg 1990; 1999).

Vote choice

We examine the final vote choice by conducting a logistic regression in which we predict the likelihood of an in-party vote choice using the standard set of predictors we have used so far: party affiliation, strength of partisanship, the initial partisan preference score, and the sex of the in-party and out-party candidates. We find in Table 4, below, that our findings are largely in line with what we would expect given our previous results.

Table 4. Logistic Regression on In-Party Vote

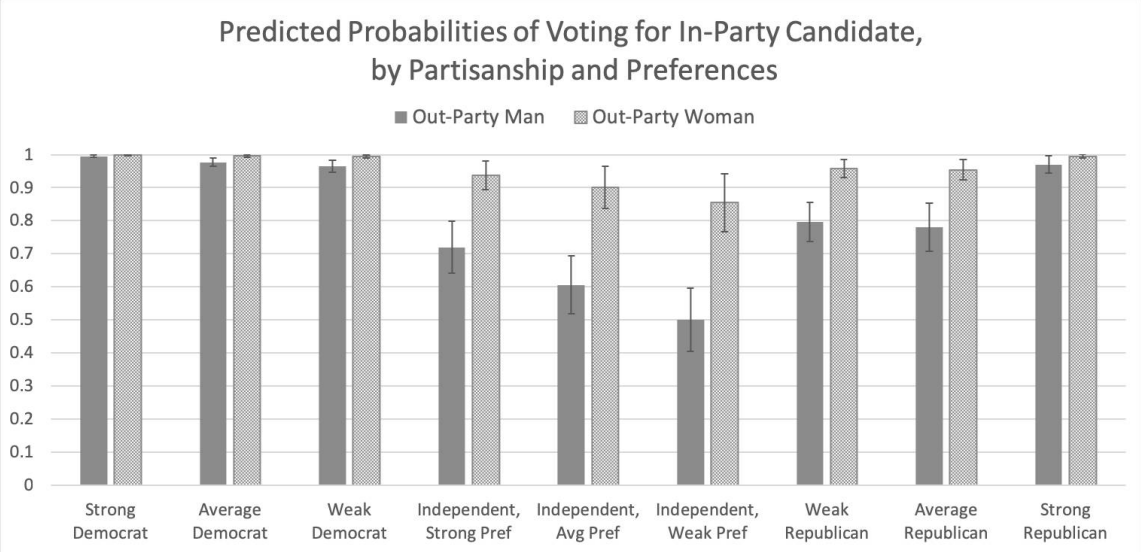
Variable	Coef (Std. Err)
Democrat	1.808** (0.561)
Strong PartyID	-0.351 (0.259)
Initial Party Preference	0.085*** (0.021)
In-Party Woman	-0.215 (0.559)
Out-Party Woman	1.774* (0.712)
Constant	0.047 (0.406)

N = 278 F Stat Sig. > 0.000 Pseudo R2 = 0.352
* - p < .050, ** - p < .010, *** - p < .001

Democrats were more likely to vote for their in-party candidate than Republicans were to vote for their in-party candidate. This is no surprise, since we have found that the Democrats in the sample consistently rated their in-party candidate higher than Republicans rated their in-party candidate. The initial party preference score is also a significant predictor, again aligning with our previous results on candidate evaluations. Subjects who started out with stronger initial preferences for one party over the other were indeed more likely to vote for their preferred party. Our key gender variables provide only one significant result in our voting model, however. The In-Party Woman variable was not significant, and did not appear to influence subject behavior when voting. It did however have a negative coefficient, which is

surprising given the generally positive findings associated with the presence of an in-party woman on preferring the in-party candidate. The Out-Party Woman variable was significant, and positive, suggesting that when the *out-party* candidate was a woman, subjects were more likely to vote for their *in-party* candidate. To find out how much more likely, we calculate predicted probabilities, plotted in Figure 8 below.

Figure 8. Predicted probabilities of voting for the in-party candidate, by Candidate Gender



We generated predicted probabilities for each position on the seven-point party ID scale, and set the Initial Partisan Preference score at the mean for each partisan group (so for example, the average Weak Democrat had an initial preference of 27 points, while an average Weak Republican had an initial preference of 20 points).[13] For Independents, we took the 25th, 50th and 75th centile score, approximating weak, moderate and strong party preferences for those who do not align with either party. The results are striking.

For Democrats, seeing an out-party man resulted in a slight decrease in the likelihood of voting for the in-party candidate from .997 for Strong Democrats to .965 for Weak Democrats (Table A9 in the

[13] For Independents we used the values at the 25th, 50th and 75th percentile of the preference distribution.

Appendix). When subjects saw an out-party woman however, the likelihood stayed above .994 regardless of strength of partisanship.

For Republicans, the results are stronger. Republicans who saw a Democratic man defected surprisingly often, voting for their in-party candidate with probabilities of about .80 for Weak and Average Republicans and .943 for Strong Republicans. But when the Democrat was a woman, in-party voting probabilities were above .950 for each group.

The strongest effects are seen for Independents. Without strong partisan ties, Independents are always more likely than true partisans to defect away from their preferred party's candidates. When confronted with an out-party man, Independents with stronger preferences still were only about 70% likely to vote for their in-party candidate, while those with moderate preferences were about 60% likely to do so. The most unaligned people in our sample – Independents with almost no preference between the parties - were about 50/50 on whether they would vote for their preferred-party candidate, as might be expected (0.501). When the out-party candidate is a woman however, it greatly encouraged Independents to stay loyal. When those most unaligned of Independents viewed a woman in their non-preferred party, they increased the likelihood of voting for the preferred-party drastically, up to about .85 – a more than 35-point increase in the probability of an in-party vote. Independents with moderate and stronger preferences similarly jumped, both becoming more than 90% likely to vote with their initial party preference. Clearly, this is an indication that the gender of candidates can and does have important influences over voting decisions, but that the effects are also closely moderated by partisan ties.

Discussion and Conclusions

In this study, we have been able to examine the effects of candidate sex in both the in- and out-party at numerous points, and in numerous ways, throughout a simulated political campaign. We find considerable support for our hypotheses. On Day 1, we find clear evidence of gender effects. Subjects differed in what information they searched for about the candidates based upon the sex of the candidates they saw and in which party those female candidates appeared. Subjects with a female candidate in their in-party sought out more information about both in- and out-party candidates, "checking up" on both in

order to help them make their vote decision. Subjects with female candidates in their out-party, though, looked for significantly less information about both candidates, suggesting that their decision was easier to make.

Subjects also subsequently evaluated those candidates differently, though the pattern of these results was slightly different than we had anticipated. We hypothesized that out-party women would lead to higher ratings for in-party (male candidates) but that in-party women would be liked less than in-party men. Our results suggest, however, that a woman in either a subject's in-party *or* out-party leads subjects to rate their in-party candidate more highly (on Day 1, at least). We do not have a compelling explanation as to why we find this, but several prior DPTE studies that examine candidate gender find similarly that subjects evaluate female candidates more highly than male, even when their actions do not necessarily correspond with their professed attitudes. It is also important to note that this finding is only statistically significant on Day 1 and washes out during the rest of the campaign.

In general, the role of gender seemed to diminish as the campaign wore on. We find in our analysis of Days 2-10 that differences in information search largely disappeared, and differences in candidate evaluation diminished in size, becoming statistically insignificant for both in- and out-party candidates as the campaign wore on. While this may be, in part, due to attrition in our sample, the largest drop-off by far was from Day 1 to Day 2, after which the sample size remained fairly consistent (and even increased again as the study entered its final few days). Even if we ignore Day 1 in our over-time analysis, the pattern of results is the same. Candidate sex largely mattered at the beginning of a campaign, when voters first met and began forming evaluations of candidates, and then at the end, when voters cast their ballots. In between, gender largely receded into the background. In other words, understanding *how* gender matters during a campaign is also a question of understanding *when* during the campaign gender is influential.

This makes sense. When subjects/voters first meet candidates, they may have few *a priori* expectations of them, but gain a few quick cues. First impressions are likely dominated by the partisanship of the candidate, as well as their name and their physical appearance, both of which can

convey gender (as well as, potentially, things like race, ethnicity, age and attractiveness) (Fiske 1998; Lau and Redlawsk 2006; Taber and Lodge 2013). From that initial introduction to a candidate, before a subject/voter learns more about them, these are the foundations upon which candidate evaluations are built. To some degree, they are "baked in" from that initial meeting and only slowly change over time. From this study, we can track that change, and demonstrate that it does in fact happen.

Our results have important implications for women candidates for office, as well. First, our findings suggest that the early part of a campaign—or at least the first several times a voter encounters a woman running for office—are particularly important. This may mean that female candidates should be very conscious of first impressions and be careful about addressing gender-stereotypes head-on early in their campaigns. Prior research suggests that competence judgments are especially important for female candidates (Ditonto 2017; Ditonto 2018), so perhaps paying particular attention to competence-related information in the earliest stages of a campaign would be beneficial. This may be particularly wise given findings by other researchers that voters take less time to make their vote decision when one candidate is a woman (Fulton and Ondercin 2013)

Second, our findings related to partisan preference suggest that women candidates should be particularly attuned to gender considerations for independent and weakly-affiliated partisans. Strong partisans do not seem to be fazed much by a candidate's sex, but those who are less attached to a political party do seem to incorporate candidate sex into their decision-making. This is an important finding both for women candidates and for scholars of gender and politics. While research has clearly shown that female candidates fare differently in the Republican vs. Democratic party (e.g. Plutzer and Zipp 1996; Schneider and Bos 2016; Hayes 2005, 2011), to our knowledge, this is the first study that has looked for and found different effects of candidate sex by *strength* of partisanship. It seems as though female candidates in the Republican party may be particularly disadvantaged among weak partisans when they run against a male Democrat and that female candidates in general may be at a disadvantage among independent voters when they face off against a man. Women who run for office may want to consider this dynamic in their campaign strategy.

Finally, women candidates may wish to consider their overall campaign strategy in light of these findings. Political campaigns often struggle to allocate resources to where they are most effective, and the choice is often framed as being between rallying supporters, persuading unaffiliated voters and converting opposing partisans. These findings suggest that women candidates may do best by focusing on rallying their own base, then turning to persuading unaffiliated voters to support her candidacy. Opposing partisans seem highly unlikely to be open to conversion, and thus an inefficient place to expend resources.

Another contribution of this study is the introduction of a new approach to the study of how voters learn about and form evaluations of candidates during a campaign. The multi-day study we conducted gave us a unique window into how subjects chose to learn about candidates, how that learning influenced their evaluations, and how, ultimately, they chose to vote based upon what they learned and how they felt. Since this is an experiment, we were able to present our candidates as either a man or woman, allowing us to isolate the effects that candidate gender had throughout the campaign. At the same time, the high-information and dynamic nature of the study more closely approximates a real-world campaign than simple, "vignette"-style experiments. This method presents researchers with an opportunity to begin "bridging the gap" between lab and survey experiments with high internal validity, and externally-valid real-world studies that do not allow for the isolation of causal mechanisms.

While we learned a great deal from this study, we feel that we have just scratched the surface of this methodological technique. A great deal of what we know about voter decision-making and candidate evaluation comes from either short experiments providing minimal information to candidates or from observational studies where voters are surveyed at one or two points during a campaign, but where researchers have no ability to determine what voters actually learned about the candidates. These two techniques have contributed the vast majority of what we know about campaigns and voters, but have still left large gaps in between telling us when stimuli *can* produce effects and which stimuli tend to actually do so. The only way to determine that is to closely track what information people actually look at and how it changes their evaluations and behaviors over time.

In the future, we hope to expand our use of this system to incorporate videos simulating campaign advertisements, journalistic articles describing the campaign to simulate the media, and of course, larger and more representative samples.

References

Andersen, David J. and Tessa M. Ditonto. 2018. Information and its presentation: Treatment effects in low- vs high-information experiments. *Political Analysis*. 26(4): 379-398. Accessible online at: https://doi.org/10.1017/pan.2018.21

Andersen, David J. and Richard Lau. 2018. Pay rates and subject performance in social science experiments using crowdsourced online samples. *Journal of Experimental Political Science.* 5(3): 217-229. Accessible online at: https://doi.org/10.1017/XPS.2018.7

Bargh, John A. 1999. The cognitive monster: the case against the controllability of automatic stereotype effects. See Chaiken & Trope 1999, pp. 361–82

Bauer, Nichole M. 2015a. Emotional, Sensitive, and Unfit for Office? Gender Stereotype Activation and Support Female Candidates. *Political Psychology*, 36(6): 691-708.

Bauer, Nichole M. 2015b."Who stereotypes female candidates? Identifying individual differences in feminine stereotype reliance." *Politics, Groups, and Identities* 3(1): 94-110.

Bauer, N.M., 2017. The effects of counterstereotypic gender strategies on candidate evaluations. *Political Psychology*, *38*(2), pp.279-295.

Brooks, Deborah Jordan. 2013. *He Runs, She Runs: Why Gender Stereotypes do not Harm Women Candidates*. Princeton, NJ: Princeton University Press.

Burrell, Barbara. 1996. *A Woman's Place is in the House: Campaigning for Congress in the Feminist Era*. University of Michigan Press.

Chaiken S. 1980. Heuristic versus systematic information processing and the use of source versus message cues in persuasion. J. Personal. Soc. Psychol. 39:752–66

Chaiken S, Trope Y, eds. 1999. Dual-Process Theories in Social Psychology. New York: Guilford.

Chong, Dennis and James N. Druckman. 2013. Counterframing effects. *Journal of Politics*. 75(1): 1-16.

Cook, Elizabeth Adell. 1994. Voter responses to women candidates. In Cook, Elizabeth Adell, Sue Thomas, and Clyde Wilcox (eds.) *The year of the women: Myths and realities.* Westview: Boulder, CO.

Darcy, R., S. Welch, and J. Clark. 1994. *Women, Elections, and Representation.* Lincoln: University of Nebraska Press.

Ditonto, Tessa, 2017. A High Bar or a Double Standard? Gender, Competence, and Information in Political Campaigns. *Political Behavior*, *39*(2), pp.301-325.

Ditonto, Tessa. 2018. A Face Fit for Office? Appearance Based Competence Inferences in High-Information Environments. *Electoral Studies*, 54: 248-253

Ditonto, Tessa, Allison Hamilton, and David Redlawsk. 2014. "Gender Stereotypes, Information Search and Voting Behavior in Political Campaigns." *Political Behavior,* 36: 335-358.

Dolan, Kathleen. 1998. Voting for Women in the "Year of the Woman." *American Journal of Political Science*, 42(1), 272-293

Dolan, Kathleen A. 2004. *Voting for Women: How the Public Evaluates Women Candidates*. Boulder: Westview Press.

Dolan, Kathleen. 2014. *When does Gender Matter? Women Candidates and Gender Stereotypes In American Elections.* New York: Oxford.

Druckman, James N., Jordan Fein and Thomas J. Leeper. 2012. A source of bias in public opinion stability. *American Political Science Review*. 106(2): 430-454.

Eagly, Alice H. and Karau, Steven J. 2002. Role congruity theory of prejudice toward female leaders. *Psychological review*, 109(3): 573-598.

Fiske, Susan T. 1998. Stereotyping, prejudice, and discrimination. In D. T. Gilbert, S. T. Fiske, & G.Lindzey (Eds.), *The handbook of social psychology* (4th ed., *Vol. 2*, pp. 357–411). Boston: McGraw-Hill.

Fiske, Susan T. and Steven L. Neuberg. 1990. A Continuum of Impression Formation, from Category-Based to Individuating Processes: Influences of Information and Motivation on Attention and Interpretation. *Advances in Experimental Social Psychology. 23: 1-74.*

Fiske, Susan T., Monica Lin, and Steven L. Neuberg. 1999. The continuum model: Ten years later. In Shelly Chaiken and Yaacov Trope (Eds.), *Dual-process Theories in Social Psychology.* New York City, New York: Guilford Press.

Fulton, Sarah and Heather L. Ondercin. 2013. Does sex encourage commitment? The impact of candidate choices on time-to-decision. *Political Behavior*. 35(4): 665-686.

Gerber, Alan S., James G. Gimpel, Donald P. Green, and Daron R. Shaw. 2011. How large and long-lasting are the persuasive effects of televised campaign ads? Results from a randomized field experiment. *American Political Science Review*. 105(1): 135-150.

Hayes, D., 2005. Candidate qualities through a partisan lens: A theory of trait ownership. *American Journal of Political Science, 49*(4), pp.908-923.

Hayes, D., 2011. When gender and party collide: Stereotyping in candidate trait attribution. *Politics & Gender, 7*(2), pp.133-165

Holman, Mirya R., Merolla, Jennifer L. and Elizabeth J. Zechmeister. 2011. Sex, stereotypes, and security: A study of the effects of terrorist threat on assessments of female leadership. *Journal of Women, Politics & Policy, 32*(3), 173-192.

Huddy, Leonie and Theresa Capelos. 2002. Gender Stereotyping and Candidate Evaluation: Good News And Bad News for Women Politicians. In *The Social Psychology of Politics*. Springer US: 29-53.

Huddy, Leonie, and Nadia Terkildsen. 1993. The Consequences of Gender Stereotypes for Women Candidates at Different Levels and Types of Office. *Political Research Quarterly*, 46(3), 503-525.

King, David C. and Richard E. Matland. 2003. Sex and the Grand Old Party: An Experimental Investigation of the Effect of Candidate Sex on Support for a Republican Candidate. *American Politics Research*, 31(6), 595-612.

Lau, Richard R. and David P. Redlawsk. 1997. Voting Correctly. *American Political Science Review,* 91(September), 585-599.

Lau, Richard R. and David P. Redlawsk. 2001. Advantages and Disadvantages of Cognitive Heuristics in Political Decision Making. *American Journal of Political Science*, 45(October), 951-971.

Lau, Richard R., and David P. Redlawsk. 2006. *How Voters Decide: Information Processing during an Election Campaign*. New York: Cambridge University Press.

Lau, R.R., Andersen, D.J., Ditonto, T.M., Kleinberg, M.S. and Redlawsk, D.P., 2017. Effect of Media Environment Diversity and Advertising Tone on Information Search, Selective Exposure, and Affective Polarization. *Political Behavior*, *39*(1), pp.231-255.

Lawless, Jennifer L. and Fox, Richard L., 2005. *It takes a candidate: Why women don't run for office*. Cambridge University Press.

Lawless, Jennifer L. & Richard L. Fox. 2010. *It Still Takes a Candidate: Why Women Don't Run for Office*. Cambridge: Cambridge University Press.

Mitchell, Dona-Gene. 2012. It's about time: The lifespan of information effects in a multiweek campaign. *American Journal of Political Science*. 56(2): 271-518.

Mitchell, Dona-Gene. 2013. In search of enduring information effects: Evidence from a panel experiment. *Electoral Studies*, 32(1), 101-112.

Mitchell, Dona-Gene. 2014. Here today, gone tomorrow? Assessing how timing and repetition of scandal information affects candidate evaluations. *Political Psychology*. 35(5), 679-700.

Petty RE and JT Cacioppo. 1981. Attitudes and Persuasions: Classical and Contemporary Approaches. Dubuque, IA: Brown

Plutzer, Eric and John F. Zipp. 1996. Identity Politics, Partisanship, and Voting for Women Candidates. *Public Opinion Quarterly*, 60, 30-57.

Sanbonmatsu, Kira. 2006. Do parties know that "women win"? Party leader beliefs about women's electoral chances. *Politics & Gender*, *2*(4), pp.431-450.

Sanbonmatsu, Kira. 2010. *Where women run: Gender and party in the American states*. University of Michigan Press

Sanbonmatsu, Kira, & Kathleen Dolan. 2009. Do Gender Stereotypes Transcend Party? *Political Research Quarterly,* 61(1) 79-89.

Schneider, M.C. and Bos, A.L., 2014. Measuring stereotypes of female politicians. *Political Psychology*, *35*(2), pp.245-266.

Schneider, M.C. and Bos, A.L., 2016. The Interplay of Candidate Party and Gender in Evaluations of Political Candidates. *Journal of Women, Politics & Policy*, *37*(3), pp.274-294.

Seltzer, Richard, Jody Newman, and Melissa Voorhees Leighton. 1997. *Sex as a Political Variable: Women as Candidates and Voters in US Elections.* Boulder, CO: Lynne Rienner.

Biographical Statements

David J. Andersen is an assistant professor of political science at Iowa State University in Ames, IA 50010.

Tessa Ditonto is an assistant professor of political science at Iowa State University in Ames, IA 50010.

Online Appendix

Experimental Recruitment, Payment and Turnout

On the first day of the study we recruited 400 subjects through Amazon's Mechanical Turk. A Mechanical Turk sample was preferable in this case for two major reasons: acceptability and practicality. First, this was the first time a study of this design had been attempted, and we did not aim to use a representative sample and make claims about how these results pertained to specific subgroups within the nation. We sought to identify general reactions to our treatment with an American voter-age sample, and a Mechanical Turk sample allowed us to do this, as would other samples.

MTurk samples have repeatedly been shown to not differ meaningfully from other types of samples (see Buhrmester, Kwang and Gosling 2011; Berinsky, Huber and Lenz 2012; Weinberg, Freese and McElhattan 2014, Clifford, Jewel and Waggoner 2015). One of the primary concerns about MTurk sample demographics is that MTurkers tend to be more liberal than nationally representative samples, and more Democratic (Berinsky, et al 2012; Huff and Tingley 2015). However, for this study a more liberal sample presents a tougher test than one that is more conservative, since conservatives are more likely to hold traditional views on gender, and Republican women tend to fare worse than Democrats (e.g. King and Matland, 2003, Dolan 2004). We anticipated having a greater share of liberal Democrats in our sample, but felt that this was only likely to make our treatment effects weaker, diminishing the likelihood of committing a Type 1 error. While a professionally-recruited representative sample would have also been desirable, our second consideration tipped the scales in favor of MTurk.

The practicality of the costs of an MTurk sample made this study feasible. We expected that with a 10-day study such as this, not all of our subjects would choose to participate each day. If we paid a flat rate for participation within the study, we would be paying for non-participation, and potentially even incentivizing non-participation. With professionally-recruited samples, the anticipated daily drop-off of subjects did not affect our costs. We recruited quotes from several recruitment firms (YouGov, SSI, and Qualtrics), but all insisted that subjects would be paid a flat rate based upon an acceptability criterion,

such as participating in a certain number of days.[1] Subjects who fell below the threshold (say, seven days) would not be paid, and those who completed more than that would be. While this was acceptable, we feared that our subjects would simply all drop out following day 7, leaving us with no voting data, which was critical to our design. Even if subjects were told that participation on day 10 was required, this now presented an incentive to skip days 7,8, and 9, presenting a different problem. The threshold payment mechanism seemed to provide incentives to our subjects that we felt were not ideal.

Mechanical Turk gave much more flexibility both in participation and in cost. We were able to structure our Mechanical Turk system so that subjects could, but did not have to, participate in each day (or not) and be paid precisely for their level of participation. We feel that this more closely approximates the "benefits" of monitoring and paying attention to normal campaigns and created no incentive to not-participate. The cost savings were also enormous.

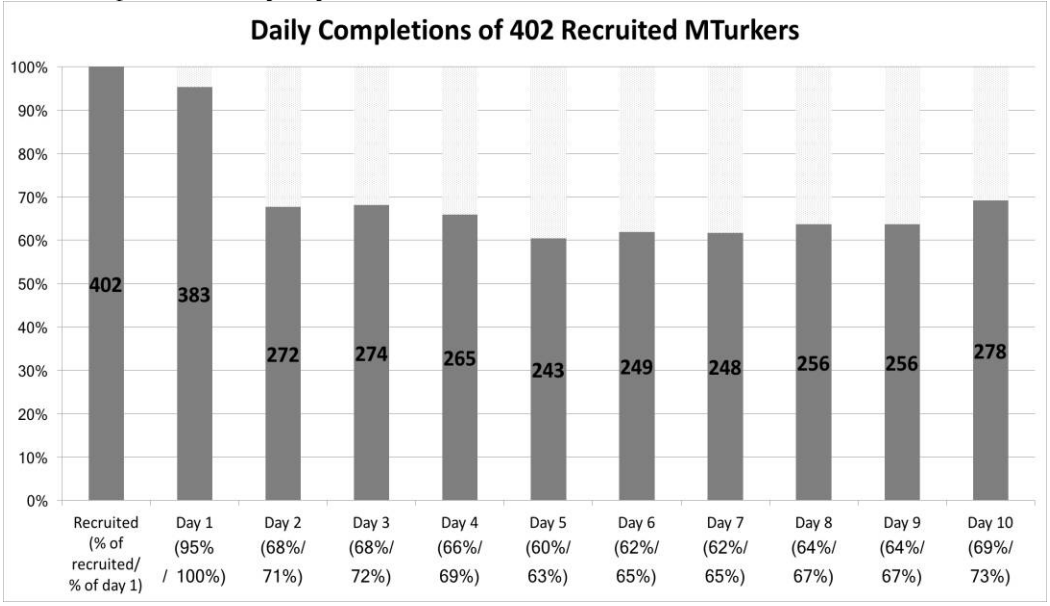
We recruited our subjects by offering payment of up to $16, a rather large sum on MTurk. The payment was tied to daily participation however. Completing day 1 immediately paid $4, but days 2 thru 9 paid only $1 each. The final day when voting would occur again paid $4. The first $4 was paid via the standard Mechanical Turk payment system, but the remaining $12 was paid as a bonus after the study concluded. Thus, subjects were incentivized to return each day, but were not forced to sign in. This mimics the experience of real campaigns, where people are free to pay attention to politics and the information around them, but are not forced to on a daily basis. They are, however, encouraged to vote (or at least more encouraged to vote than to pay attention on a daily basis).

We had to have sufficient funds available to pay all of our subjects, in case they all participated on each day. With our payment system, this meant we had to have $8000 set aside (($4 * 400 subjects + MTurk's 40% standard commission) + ($12*400 subjects + MTurk's 20% commission for individual

[1] One of our quotes also directly stated that the company did not like the idea of running a panel in this way, because the questions asked in it could be used to generate rival data to their own tracking polls on politics. They indicated that they were willing, but not eager to supply a sample, and their quote was predictably not competitive.

bonuses) For comparison, our closest quote from a professional organization was for approximately $12,000. Beyond the immediate savings in cost, we additionally "saved" money each time a participant failed to return to the study. While it is not desirable to lose participants, this was at least efficient. In total given our turnout, our costs for the study amounted to only $6,050.

Figure A1. Participation rate by day



From the 400 subjects who were recruited the first day, 383 successfully completed all the elements of that day's study including returning their subject ID number to us so we could track their participation. Then there was an initial drop-off in participation on day 2, where 30% of subjects failed to return. Following that decline, participation was fairly stable at about 250 subjects per day. Figure 1, above, shows the return rate for each day of study.

We have no reason to suspect that the voluntary return process created any meaningful differences in our sample or subject behavior and there is no evidence that non-returners varied systematically from those who chose to continue with the study. While turnout on each day fluctuated between 65% and 73% of the number of completed subjects from day 1, there were no demographic differences day-to-day. That is, when running t-tests to see whether the average participant on each subsequent day differed from the average subject on day 1, we found no significant differences.

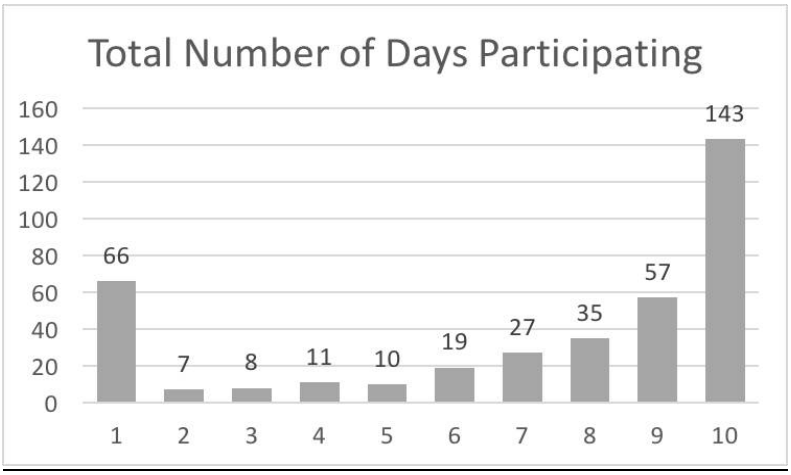
Table A1, below, shows the daily percentages of the sample that identified as female or black, and the average age, partisanship and the initial in-party preference scores of the sample. From this, we can rule out some likely factors that might affect the likelihood of subjects returning, such as strong partisans either being more or less interested in learning more about the candidates over time, or men encountering women candidates and altering their interest in the campaigns. We simply find no large differences in the composition of the sample over time. To check for this, we conducted t-tests between the daily average composition of the sample by various demographic factors and the composition of the sample on day 1. We found no significant differences on any measure, on any day.

Table A1. Daily demographics of the sample

	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9	Day 10
N	383	272	274	265	243	249	248	256	256	278
% Female	43.6	43.0	44.2	42.3	44.9	41.0	44.0	42.2	43.4	44.2
% Black	5.0	3.7	4.4	4.2	4.5	4.4	4.8	5.1	3.9	4.0
Age	33.6	34.4	34.0	34.1	34.2	34.3	33.8	34.1	34.3	34.4
PartyID	3.12	3.15	3.08	3.08	3.08	3.13	3.15	3.09	3.12	3.12
Init Pref	33.30	33.18	34.70	34.38	35.38	33.55	33.79	33.84	33.95	34.03
* - sig. >.050, t-tests compared to day 1 results										

Over the course of the 10-day study, a large plurality of our subjects (143) returned every day, and turnout overall skewed towards participating more rather than less. The average number of days that a subject participated was 7.1 out of the 10, showing that subjects were diligent in returning day after day. The distribution is bimodal however, with the second largest number of days participated at 1. These were our subjects who completed the study on day 1, but never returned. We can only guess at why these subjects failed to return, but the likely reasons are that they either lost or incorrectly saved their hyperlink to the study, or they simply did not understand the directions they were provided. Since the people who never returned to the study did not differ in any measurable ways from the people who did return, we have no reason to suspect that they biased the sample.

Figure A2. The number of days subjects participated in the study



Another concern is that the balance between the groups in the study (Two Men; Democratic Woman; Republican Woman) could have been off initially or through attrition. To check this, we conducted oneway ANOVA's looking for statistical differences in the rate at which these demographic groups appeared in either the control (Two Men) or either treatment group. We examined day 1 (initial randomization), day 2 (following the initial dropoff of subjects), and day 5 (the lowest day of participation). Below, in tables A2, A3, and A4 we report the f-statistic significance levels for those ANOVAs. None cross the .050 level of statistical significance.

Table A2. Balance checks on treatments, day 1

N=383	Female	Black	Age	Democrat	Republican
Woman Democrat	0.648	0.996	0.797	0.652	0.282
Woman Republican	0.395	0.833	0.954	0.191	0.491

Table A3. Balance checks on treatments, day 2

N=272	Female	Black	Age	Democrat	Republican
Woman Democrat	0.560	0.646	0.815	0.292	0.245
Woman Republican	0.833	0.063	0.633	0.106	0.452

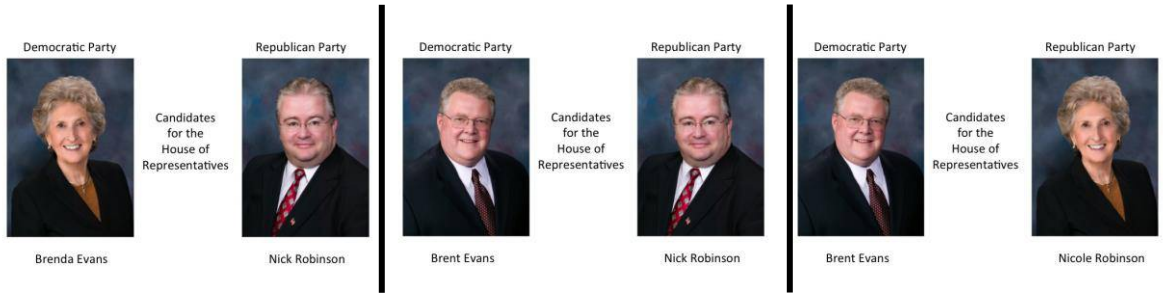
Table A4. Balance checks on treatments, day 5

N=243	Female	Black	Age	Democrat	Republican
Woman Democrat	0.873	0.341	0.324	0.091	0.111
Woman Republican	0.909	0.164	0.851	0.067	0.148

Daily Procedures

On the first day of the study subjects followed a link to sign into the DPTE system, where they first completed a 65-item pre-questionnaire that asked about their political opinions, participation and knowledge, demographics and other items. Then then participated in a brief 2-minute simulation that introduced them to the DPTE system. Once familiar with the system, they were introduced to the campaign by first viewing a picture a side-by-side picture of the two candidates (Figure A3, below) and participating in another short 3-minute dynamic information board that provided basic background information about the two candidates. Finally, subjects were asked to evaluate the candidates, and saw them for the first time. Subjects (unbeknownst to them) were first randomized to either see a woman candidate or not, and then if they were to see a woman candidate, to see her either as a Republican or a Democrat. This, unfortunately gave us an over-large control group whom only saw men candidates. The pictures we used were pre-tested for comparability on attractiveness, competence, age and likeability.

Figure A3. Candidate pairings



At the end of the day's session, subjects were then provided a link to return to the study each day for the next 9 weekdays. That link contained an embedded identification code that allowed us to track subject participation each day and maintained the gender manipulation throughout the duration of the study. To return to the study, they simply had to click on that link (or paste it into their browser window) once per day. Each day that subjects returned, they participated in an approximately 3-minute long dynamic information board that presented them with new information about the candidates.

On the first day, subjects were presented with five attributes about each candidate, all of which related to their background/demographic information. On each subsequent day, two new items appeared alongside

two current event news reports that created a theme for the day. These items were presented as boxes that scrolled down the computer screen. Each box contained a description of the information it contained (for example, "Robinson's views on Abortion;" "Evans' Political Experience;" or "Associated Press Breaking News Report on Terrorism"). Subjects were able to view the information items by clicking on the scrolling box, which would then open and present the information inside to be read. Each day presented approximately 6 political items about each candidate to subjects, but they were free to open as many, or as few, as they wished. Each item could appear multiple times, and overall subjects could choose between about 40 different boxes on any given day[2]. Figure A4 (also in the main text as Figure 1), below, shows the universe of information subjects could view, and when and how often it was available during the study.

Figure A4. Political Information Items available each day

Attributes	Availability									
	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9	Day 10
Abortion Policy	0	4	2	1	1	0	0	0	1	1
Iran Policy	0	0	0	0	0	0	0	4	2	1
Crime Policy	0	0	0	0	0	0	0	0	4	1
Economic Philosophy	0	0	4	2	1	1	0	0	0	1
Editorial About	0	0	0	4	2	1	1	0	0	1
Education	2	1	0	0	0	0	0	0	0	1
Education Policy	0	0	0	0	0	4	2	1	1	1
Energy Policy	0	0	0	0	4	2	1	1	0	1
Family	2	1	0	1	0	0	0	0	0	1
Global Warming Policy	0	0	0	0	0	4	2	1	1	1
Gun Control Policy	0	4	2	1	1	0	0	0	1	1
Healthcare Policy	0	0	0	0	0	0	4	2	1	1
Immigration	0	0	0	0	0	0	4	2	1	1
Jobs Policy	0	0	0	4	2	1	1	0	0	1
Political Experience	3	2	1	0	0	0	0	0	0	1
Religion	3	1	1	0	0	0	0	0	0	1
Tax Policy	0	0	4	2	1	1	0	0	1	1
Terrorism Policy	0	0	0	0	4	0	0	0	0	0
Social Philosophy	3	2	1	0	0	0	0	0	1	1
Defense Policy	0	0	0	0	0	0	0	4	2	1
Items per candidate	13	15	15	15	16	14	15	15	16	19
Total	26	30	30	30	32	28	30	30	32	38

[2] Information items provided descriptions of the candidates' stances on policy issues, as well as various aspects of their background and family. All information was presented in a neutral tone and mimics the sort of information available on candidate websites. The full text of the information items can be seen by viewing the completed study on the DPTE website (see footnote 3).

Information items typically appeared on multiple days, but varied in how many times they were presented on each day. On the first day, new items would appear four times for each candidate, emphasizing their presence. On subsequent days, the frequency of their appearance tapered off, matching the typical rise and decline of information about candidates during a campaign news cycle.

At the end of each day's session, subjects were asked to evaluate both candidates on a feeling thermometer. Finally, on the tenth day, we presented no new information, but instead made each information item that had been presented in the study appear again in the dynamic information board two more times. Subjects then proceeded to vote, and evaluate the candidates on feeling thermometers again.

Information Search

Subjects were free to interact with the dynamic information boards in any way they wished, and could view as much or as little information as they wanted about the candidates given what was available. Given our theory of information processing, we would expect subjects who were less certain about their candidate preferences – and thus who they should vote for – would be likely to seek out more information about the candidates. This would provide them with more information with which to form a good opinion of the candidates. Seeing a woman candidate in their preferred party – when candidates are expected to be men – could create just this kind of uncertainty. Becoming more certain that their preferred party's candidate is the better option would reverse this, leading to a lower information search. Seeing a woman candidate in the out-party might provide that increased certainty.

Table A5. Information items viewed on day 1, by candidate gender

	Day 1 Total Search* (SE)	Day 1 In-Party Search* (SE)	Day 1 Out-Party Search (SE)
Out-Party Woman (105)	11.01 (4.77)	5.66 (2.45)	5.35 (2.72)
Two Men (189)	11.67 (4.72)	5.92 (2.63)	5.76 (2.45)
In-Party Woman (89)	12.73 (4.64)	6.54 (2.52)	6.19 (2.43)
sig.	0.040*	0.050*	0.071

On day 1, this is what we observe in the results (Figure 1 and Table A5, above). Subjects always look for more information about their in-party candidate than their out-party candidate, but they look for more information for both when there is an in-party woman, and less for both when there is an out-party woman. These are statistically significant differences on for the In-Party search and the Total search, but falls just shy of convention standards for the Out-Party Search. We expect that the differences in information search will be greatest on day 1, and tend to decrease over time, because on the first day gender cues will be most prominent. As subjects learn actual information about the candidates, they should rely less upon these cues.

Table A6. Information search total and total by candidate, by candidate gender

N= 383	Total Search (SE)	Total Candidate	
		In-Party (SE)	Out-Party (SE)
Out-Party Woman	9.56 (0.47)	4.76 (0.23)	4.57 (0.24)
Two Men	9.94 (0.28)	5.03 (0.16)	4.75 (0.14)
In-Party Woman	10.36 (0.43)	5.18 (0.22)	5.01 (0.21)
sig.	0.414	0.386	0.344

We can examine this by aggregating all of the days' information searches together, looking for differences in search throughout the entire campaign. In Table A6, above (and the associated Figure 4) we find an identical pattern in the results found on day 1, but the statistical significance has washed out. This is in-line with our expectations. The role played by candidate gender cues does not completely go away, but should tend to erode over time. Depending on the size of the sample used, statistical significance should persist for varying rates of time, but the general pattern should be present in the data. We can examine this by looking at periods within the campaign, to see if differences in information search are apparent at any given section: Early (days 1-3), Middle (days 4-6), or late (days 7-9). Here we use the average number of items viewed on days during each period, and find the pattern we expect to see, though no statistical significance.

Table A7. Periods of candidate information search, by candidate gender

	Early (383)		Mid (272)		Late (243)	
Candidates	In-Party (SE)	Out-Party (SE)	In-Party (SE)	Out-Party (SE)	In-Party (SE)	Out-Party (SE)
Out-Party Woman	5.32 (0.20)	5.09 (0.22)	5.04 (0.33)	4.80 (0.31)	4.05 (0.27)	3.86 (0.29)
Two Men	5.68 (0.16)	5.39 (0.15)	4.80 (0.16)	4.72 (0.16)	4.11 (0.17)	3.86 (0.15)
In-Party Woman	6.02 (0.24)	5.71 (0.23)	4.84 (0.24)	4.79 (0.22)	3.85 (0.23)	3.84 (0.25)
sig.	0.094	0.134	0.744	0.961	0.721	0.998

Again, in every period subjects examined more information for their in-party candidate than the out-party candidate. Initially, in the early stage of the campaign subject maintain a pattern of looking at more information for the candidates when the in-party candidate is a woman, and less information about them when the out-party candidate is a woman. But this pattern fades and even changes direction in later periods. However, no stable pattern appears in the results between the Middle and Late period, and the relatively high overlap in the confidence intervals suggests that after the first few days, subjects largely did not rely upon gender cues in determining their information search strategies. We can continue to extend this analysis down to the level of the individual day. Table A8, below, presents the results of a day-by-day analysis of the information search, split into the first week (days 1-5) and second week (days 6-10).

This information is also visualized below in Figure A5. The results show that any consistent gender effects tend to wash out of the data past about day 3. The clear pattern in information search difference on day 1 reappears on day 3, but then fades away almost completely. While a large sample could potentially turn some of these results significant, it would have to be enormous, and beyond what any normal study produces. This again supports our theory that gender cues are most influential at the outset of a campaign, when voters have little real information about candidates and are instead dominated by the cues and stereotypes that they possess.

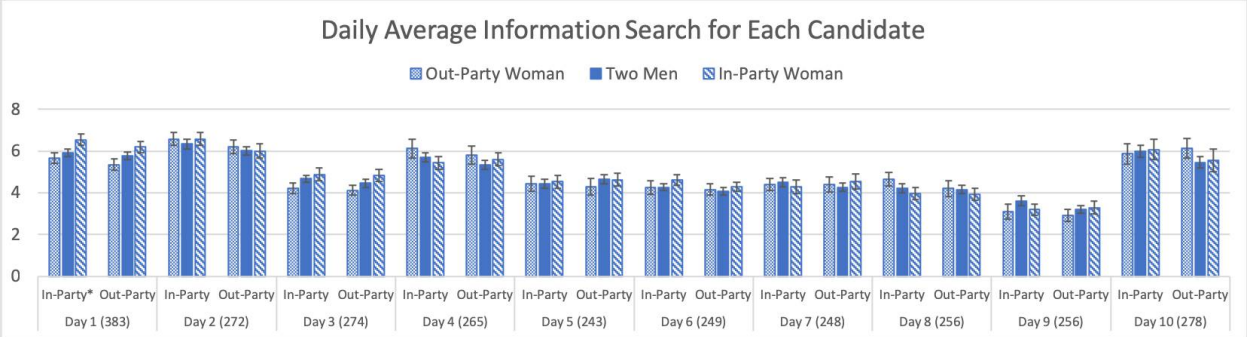
Table A8a. Week 1 information search table

	Day 1 (383)		Day 2 (272)		Day 3 (274)		Day 4 (265)		Day 5 (243)	
	In-Party* (SE)	Out-Party (SE)	In-Party (SE)	Out-Party (SE)	In-Party (SE)	Out-Party (SE)	In-Party (SE)	Out-Party (SE)	In-Party (SE)	Out-Party (SE)
Out-Party Woman	5.66 (0.24)	5.35 (0.27)	6.57 (0.31)	6.20 (0.32)	4.22 (0.26)	4.12 (0.24)	6.12 (0.46)	5.82 (0.43)	4.44 (0.36)	4.29 (0.4)
Two Men	5.92 (0.19)	5.76 (0.18)	6.33 (0.22)	6.01 (0.19)	4.67 (0.17)	4.46 (0.19)	5.69 (0.21)	5.34 (0.22)	4.43 (0.23)	4.64 (0.22)
In-Party Woman	6.54 (0.27)	6.19 (0.26)	6.58 (0.32)	6.00 (0.33)	4.88 (0.32)	4.83 (0.29)	5.43 (0.31)	5.60 (0.3)	4.53 (0.3)	4.63 (0.3)
sig.	0.050	0.071	0.727	0.851	0.181	0.171	0.384	0.526	0.973	0.664

Table A8b. Week 2 information search table

	Day 6 (249)		Day 7 (248)		Day 8 (256)		Day 9 (256)		Day 10 (278)	
	In-Party (SE)	Out-Party (SE)	In-Party (SE)	Out-Party (SE)	In-Party (SE)	Out-Party (SE)	In-Party (SE)	Out-Party (SE)	In-Party (SE)	Out-Party (SE)
Out-Party Woman	4.25 (0.34)	4.16 (0.28)	4.39 (0.30)	4.41 (0.36)	4.65 (0.34)	4.20 (0.37)	3.09 (0.35)	2.91 (0.30)	5.86 (0.47)	6.12 (0.47)
Two Men	4.26 (0.17)	4.08 (0.18)	4.51 (0.22)	4.26 (0.20)	4.23 (0.22)	4.15 (0.20)	3.62 (0.22)	3.20 (0.19)	5.99 (0.28)	5.46 (0.28)
In-Party Woman	4.61 (0.24)	4.29 (0.22)	4.29 (0.32)	4.54 (0.35)	3.96 (0.30)	3.93 (0.30)	3.19 (0.27)	3.28 (0.31)	6.07 (0.48)	5.55 (0.53)
sig.	0.570	0.815	0.848	0.777	0.300	0.813	0.311	0.601	0.943	0.426

Figure A5. Daily Average Information Search



The final, most fine-grained analysis we can produce for candidate gender effects on information search is to look issue-by-issue for gender-based differences in how subjects learned about the candidates. Besides being attracted to learning more or less information about candidates, voters may also be attracted

to learning certain types of information about women candidates rather than men candidates. Women candidates are often stereotyped as being better at some issues – education and healthcare for example – while men are deemed better at more masculine issues – such as the military and economy. Each of the 20 attributes for the candidates first appeared on a specific day, allowing us to test whether subjects were more likely to view an issue item for a candidate by a simple crosstabulation. Table A9, below, shows the Chi Square significance levels for every issue available for the two candidates, and is separated by whether the in-party or out-party candidate's gender was manipulated (we report actual values for significant findings below).

Table A9. Chi-Square differences in issue information search, by Candidate Gender

		In-Party Candidate Manipulated		**Out-Party Candidate Manipulated**	
Day	Attribute	In-Party	Out-Party	In-Party	Out-Party
1	**Education**	**0.012***	0.365	0.353	0.490
	Family	0.401	0.626	0.117	**0.008****
	Pol Exp	0.996	0.626	0.712	0.756
	Religion	0.855	0.661	0.482	0.516
	Soc Phil	0.690	0.277	0.664	0.413
2	**Abortion**	0.485	0.938	**0.014***	0.202
	Guns	0.073	0.134	0.929	0.381
3	**Econ Phil**	0.296	0.627	0.218	**0.032***
	Taxes	0.413	0.923	0.550	0.156
4	Editorial	0.478	0.597	0.530	0.407
	Jobs	0.844	0.708	0.765	0.819
5	Energy	0.566	0.240	0.852	0.796
	Terror	0.500	0.678	0.918	0.302
6	**Educ Policy**	**0.022***	0.244	0.270	0.378
	Glob Warm	0.781	0.632	0.987	0.855
7	**Healthcare**	0.137	0.199	**0.009****	0.896
	Immigration	0.448	0.137	0.783	0.933
8	Attack Ad	0.951	0.951	0.483	0.483
	Defense	**0.026***	0.494	0.330	0.452
	Iran	0.425	0.221	0.055	0.782
9	Crime	0.150	0.249	0.686	0.743

* - p < .050, ** - p < .010, *** - p < .001

The most obvious result Table A9 portrays is that there were few significant differences in which candidate attributes were viewed based upon gender differences. Given that we ran 40 tests, we would expect that out of sheer chance we would find about two significant findings given a 95% confidence level. We in fact find seven, suggesting that perhaps gender did play a role in what people chose to view about the candidates.[3] The exact issues that produce significant results are also indicative. They include Abortion, Defense Policy, Economic Philosophy, Education, Education Policy, Family, and Healthcare. All of these are issues typically discussed in the candidate evaluation literature as "gendered issues," giving us confidence that these differences are in fact related to our gender manipulation.

Table A10a. Crosstabulation of Information Search by In-Party Gender, Significant Findings

	In-Party Views				**Out-Party Views**			
	Attribute	IP Man (282)	IP Woman (101)	Chi Sq Sig	Attribute	IP Man (290)	IP Woman (93)	Chi Sq Sig
Day 1	Education	72%	86%	.012				
Day 6	Educ Pol	83%	96%	.022				
Day 8	Defense	87%	75%	.026				

Table A10b. Crosstabulation of Information Search by Out-Party Gender, Significant Findings

	In-Party				**Out-Party**			
	Attribute	OP Man (282)	OP Woman (101)	Chi Sq	Attribute	OP Man (290)	OP Woman (93)	Chi Sq
Day 1					Family	72%	57%	.008
Day 2	Abortion	97%	88%	.014				
Day 3					Econ Phil	89%	78%	.032
Day 7	Health	83%	97%	.009				

Tables A10a and A10b present the actual crosstabulation results for each of the significant findings. Beyond noting that the differences are significant however, there is little more that we can discern from these results. Since we have nothing to compare it to, and issue presentation was not

[3] We do not calculate Holms-Bonferroni corrections for multiple hypothesis testing here and instead present the raw p values. Our interests in this analysis is in detecting general patterns rather than relying strictly on statistical significance.

randomized, we can only speculate about whether it is the timing of issue appearance that led subjects to produce significant differences (four of the significant findings occur in the first three days, when cues had the strongest effects) or the issue content themselves (all of the issues can be classified as "gendered"). We present these results primarily because we think readers will be interested at seeing how information search operated at its most micro level.

Candidate Evaluation

Within the main text we present the day 1 feeling thermometer results, and present the associated table here, as Table A11. It shows a significantly higher score for the in-party candidate when either the in-party or out-party candidate is a woman, and a significantly lower score for the out-party candidate when either candidate is a woman. This in turn leads to significantly higher preference scores for the in-party candidate when there is a woman on the ballot.

Table A11. Candidate evaluations at the end of day 1, using oneway ANOVAs

	In-Party FT* (SE)	Out-Party FT* (SE)	In-Party Preference* (SE)
Out-Party Woman (105)	67.74 (17.69)	31.27 (20.73)	36.48 (28.35)
Two Men (189)	62.28 (18.75)	36.66 (20.54)	25.62 (31.56)
In-Party Woman (89)	67.67 (19.47)	31.82 (20.25)	35.85 (34.50)
sig.	0.018	0.050	0.005

Voting

While the logistic regression used in the paper allows us to control for relevant factors, the simplest way to explore if candidate gender influenced the vote choice is to generate a simple crosstabulation of voting results based upon candidate gender. The results show a clear pattern suggesting that the gender of candidates had an influence on voting patterns in a way that support our theory. Table A12, below, shows that the in-party candidate did best when appearing as a man facing an out-party woman. In that scenario, the in-party candidate received 97% of their partisans' votes. When two men

faced off against each other, there was a higher defection rate, and in-party candidates only received about 92% of the in-party vote. Worst was when an in-party woman appeared. While not staggeringly low, and not quite statistically significant, only 90% of the candidate's in-party supporters cast their vote for her. In real world elections, if this trend was present it could very well determine the fate of a competitive contest.

Table A12. Crosstabulation of In-Party voting by Candidates' Genders

	Out-Party Woman	Two Men	In-Party Woman
Voted In-Party (249)	96.30% (78)	86.52% (122)	87.50% (49)
Voted Out-Party (29)	3.70% (3)	13.48% (19)	12.50% (7)
Chi Sq. 5.578; sig. 0.061			

The logit presented in Table 4 produced predicted probabilities for the in-party vote choice as demonstrated below, in Table A13 (and the associated Figure 8).

Table A13. Predicted Probabilities of the In-Party Vote, by Out-Party Candidate Gender

	Strong Dem	Avg Dem	Weak Dem	Indep. Str. Prefs	Indep Avg. Prefs	Indep Weak Prefs	Weak Rep	Avg Rep	Strong Rep
Out-Party Man	0.995 (0.004)	0.977 (0.013)	0.965 (0.018)	0.72 (0.079)	0.606 (0.088)	0.501 (0.096)	0.796 (0.059)	0.781 (0.073)	0.97 (0.026)
Out-Party Woman	0.999 (0.001)	0.996 (0.004)	0.994 (0.005)	0.938 (0.043)	0.901 (0.064)	0.855 (0.088)	0.958 (0.028)	0.954 (0.031)	0.995 (0.006)

References

Alexander, Deborah and Kristi Andersen.1993. Gender as a Factor in the Attribution of Leadership Traits. *Political Research Quarterly* 46 (3): 527-545.

Berinsky, Adam, Gregory Huber, and Gabriel Lenz. 2012. Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis*. 20: 351-368

Buhrmester, Michael, Tracy Kwang, and Samuel D. Gosling. 2011. "Amazon's Mechanical Turk:

A new source of inexpensive, yet high-quality, data?." *Perspectives on psychological science* 6 (1): 3-5.

Clifford, Scott, Ryan M. Jewell, and Philip D. Waggoner. 2015. "Are samples drawn from Mechanical Turk valid for research on political ideology?." *Research & Politics* 2(4): 2053168015622072

Dolan, Kathleen. 2004. *Voting for Women How the Public Evaluates Women Candidates.* Boulder, CO: Westview Press.

Huff, Connor, and Dustin Tingley. 2015. "“Who are these people?” Evaluating the demographic characteristics and political preferences of MTurk survey respondents." *Research & Politics* 2(3): 2053168015604648.

King, David C., and Richard E. Matland. 2003. "Sex and the grand old party: An experimental investigation of the effect of candidate sex on support for a Republican candidate." *American Politics Research* 31(6): 595-612.

Weinberg, Jill D., Jeremy Freese, and David McElhattan. 2014. Comparing Data Characteristics And Results of an Online Ractorial Survey Between a Population-Based and a Crowdsource-Recruited Sample. *Sociological Science*, 1, 292-310.